REPORT DOCUMENTATION PAGE AD A249 262

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	March 17, 1992	

4. TITLE AND SUBTITLE

Scanning Tunneling Microscopy

5. FUNDING NUMBERS

6. AUTHOR(S)

John D. Dow

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Notre Dame
Notre Dame, Indiana 46556

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Scanning tunneling microscopy (STM) has been used to image and modify the surfaces of III-V, II-VI and group IV semiconductors. A tip-simulator based on a photodcode was developed. The simulator allows the development of ultra-sensitive electronics for controlling STM tip movement. Various forms of "nano-maching," including chiselling, sanding, and sweeping of atoms on a surface, were developed. An STM design was modified to allow bending of long thin samples of Si(100) in UHV to permit the study of surfact strain. A variety of studies were conducted on Au(in air) CdTe (in air), $Hg_{1-x}Mn_xTe$ (under glycerin), and $Hg_{1-x}Cd_xTe$ (in air and under glycerin).

14. SUBJECT TERMS

STM Images, InSb, InAs, InP, HgCdTe and Diluted Magnetic Semiconductors

15. NUMBER OF PAGES

130

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

SCANNING TUNNELING MICROSCOPY OF III-V SEMICONDUCTORS

FINAL REPORT

John D. Dow

10 March 1992

U. S. ARMY RESEARCH OFFICE

CONTRACT NUMBER DAAL03-87-K-0112

DEPARTMENT OF PHYSICS
UNIVERSITY OF NOTRE DAME
NOTRE DAME, INDIANA 46556

THE VIEWS, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE OF THE AUTHOR AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF ARMY POSITION, POLICY, OR DECISION, UNLESS SO DESIGNATED BY OTHER DOCUMENTATION.

Report

During the course of this project, our group has accomplished the following "firsts": (1) The first scanning tunneling microscopy (STM) images of the InSb, InAs, and InP (110) surfaces. (2) The first STM images of $Hg_{1-x}Cd_xTe$ surfaces. (3) The first STM images of a diluted magnetic semiconductor, $Cd_{1-x}Mn_xTe$. (4) The first demonstration that horribly convolved tips mechanically still give very good STM images. (5) The first nano-machining of Au, $Hg_{1-x}Cd_xTe$, and $Cd_{1-x}Mn_xTe$ surfaces. (5) Formation, without physical contact, of 9Å dots at room temperature on the InSb (110) surface and formation of more than one dot in a line (pixels) on that surface. (6) Observation of filling in, by diffusion and other mechanisms, of indentations made on the surface by an STM tip on HgCdTe and CdMnTe. (7) Scribing of adjacent parallel lines on the HgCdTe and CdMnTe surfaces. (8) Identification of a new c(4×6) reconstruction on some InSb (110) surfaces; this reconstruction appears to be metastable and cleavage induced. (9) Studies of the Si (100) surface under tensile compression and strain. This led to single-atom high steps moving in response to strain. (10) Observation and identification of defects and defect structures in STM images, together with theoretical description of those images.

These new results have been achieved while we have been developing new STMs with unprecedented vibrational immunity and resolution.

Publications

138. D. V. Froelich, M. E. Lapeyre, J. D. Dow, and R. E. Allen. Dependence of the GaAs (110) surface electronic state dispersion curves on the surface relaxation angle. Superlatt. Microstruct. 1, 87-89 (1985).

146. C. S. Lent, M. A. Bowen, J. D. Dow, R. S. Allgaier, O. F. Sankey, and E. S. Ho. Relativistic empirical tight-binding theory of the energy bands of GeTe, SnTe, PbTe, PbSe, PbS, and their alloys. Superlatt. Microstruct. 2, 491-499 (1986).

162. O. F. Sankey and J. D. Dow. Theory of electron spin resonance measurements of chalcogen pairs in Si. Solid State Commun. 51, 705-708 (1984).

163. W. L. Johnson, O. F. Sankey, and J. D. Dow. Deep levels associated with impurities at the bond-centered interstitial site in Si. Phys. Rev. B 30, 2070-2073 (1984).

167. J. D. Dow, R. E. Allen, and O. F. Sankey. Intrinsic and extrinsic surface electronic states of semiconductors. Chem. and Phys. of Solid Surfaces, V, 483-500 (1984), ed. by R. Vanselow and R. Howe, Springer Series in Chemical Physics 35, (Springer-Verlag, Berlin, Heidelberg, New York, Tokyo).

171. O. F. Sankey, R. E. Allen, and J. D. Dow. Theory of Si/transition-metal silicide Schottky barriers. J. Ultramicroscopy 14, 127-130 (1984).

172. O. F. Sankey, R. E. Allen, and J. D. Dow. Theory of Schottky barrier formation for transition metals on Si, diamond, Ge, and Si_xGe_{1-x} alloys. J.

Vac. Sci. Technol. B2, 491-495 (1984).

176. S. Lee, J. Sanchez-Dehesa, and J. D. Dow. Theoretical investigation of the pressure dependences of energy gaps in semiconductors. Phys. Rev. B 32, 1152-1155 (1985); B 33, 7309 (1986).

177. S. Lee, J. D. Dow, and O. F. Sankey. Theory of charge-state splittings of deep levels. Phys. Rev. B 31, 3910-3914 (1985).

178. J. D. Dow. Localized perturbations in semiconductors. In Highlights of Condensed-Matter Theory (Proceedings of the International School of Physics "Enrico Fermi", Course 89, Varenna, 1983), ed. by F. Bassani, F. Fumi, and M. P. Tosi (Societa Italiana di Fisica, Bologna, Italy, and North Holland, Amsterdam, 1985), pp. 465-494.

183. D. W. Jenkins, K. E. Newman, and J. D. Dow. Predicted energy band gaps of $(A^{III}B^{V})_{1-x}(X^{IV})_{2x}$ metastable, substitutional, crystalline alloys. Phys. Rev. B 32, 4034-4041 (1985).

185. J. D. Dow, O. F. Sankey, and R. E. Allen. Interfacial deep levels responsible for Schottky barrier formation at semiconductor/metal contacts. Appl. Surf. Sci. 22/23, 937-947 (1985).

211. R. Nicolaides, Y. Liang, W. E. Packard, Z.-W. Fu, H. A. Blackstead, K. K. Chin, J. D. Dow, J. K. Furdyna, W. M. Hu, R. C. Jaklevic, W. J. Kaiser, A. R. Pelton, M. V. Zeller, and J. Bellina, Jr. Scanning tunneling microscope tip structures. J. Vac. Sci. Technol., A6, 445-447 (1988).

229. W. E. Packard, Y. Liang, N. Dai, J. D. Dow, R. Nicolaides, R. C. Jaklevic, and W. J. Kaiser. Nano-machining of gold and semiconductor surfaces. J. Microscopy, 152, 715-725 (1988).

253. S. Y. Ren, J. Shen, R.-D. Hong, S. Klemm, M.-H. Tsai, and J. D. Dow. Deep levels and shallow-deep transitions in $ZnSe/Zn_{1-x}Mn_xSe$ superlattices. Surf. Sci. 228, 49-52 (1990).

256. W. E. Packard, Y. Liang, N. Dai, J. D. Dow, R. Nicolaides, R. C. Jaklevic, and W. J. Kaiser. Nano-machining of gold and semiconductor surfaces. Scanning Tunnelling Microscopy, ed. by W. M. Stobbs, (Blackwell, Oxford, 1988), pp. 715-725.

260. W. E. Packard, N. Dai, J. D. Dow, R. C. Jaklevic, W. J. Kaiser, and S.-L. Tang. Externally strained Si (100) observed with scanning tunneling microscopy. J. Vac. Sci. Technol. A 8, 3512-3515 (1990).

266. Y. Liang, W. E. Packard, and J. D. Dow. Fabrication of quantum dots on the InSb (110) surface. Submitted.

274. Y. Liang, W. E. Packard, and J. D. Dow. Scanning tunneling microscopy study of the cleaved InSb (110) surface. J. Vac. Sci. Technol. B9, 730-734 (1991).

DEPENDENCE OF THE GaAs (110) SURFACE ELECTRONIC STATE DISPERSION CURVES ON THE SURFACE RELAXATION ANGLE

David V. Froelich, Mary E. Lapeyre, and John D. Dow
Department of Physics, University of Notre Dame
Notre Dame, Indiana 46556
and
Roland E. Allen
Department of Physics, Texas A&M University
College Station, Texas 77843

(Received 16 July 1984)

The surface state dispersion curves $E(\bar{k})$ of the dangling bond states near the fundamental band gap, C_3 and A_5, are computed for both the established $\theta=27^\circ$ model and the recently proposed $\theta=7^\circ$ model of the (110) surface relaxation of GaAs, where θ is the surface bond rotation angle. The two models produce surface state dispersion curves that are similar to one another and to the data.

Until recently it was thought that the geometrical structure of the (110) surface of GaAs was one of the few semiconductor surface structures that was established. The accepted model was the 27° rotation model [1,2]: To a good approximation, the anions rotate rigidly out of the surface through an angle of $\theta=27^\circ$. This model was established as a result of careful analyses of low-energy electron diffraction (LEED) data, and, in addition, provided a way out of a theoretical dilemma: calculations of GaAs surface states for unrelaxed surfaces, $\theta=0^\circ$, produced surface states in the fundamental band gap (contrary to data) that receded into the valence and conduction bands when the $\theta=27^\circ$ relaxation was accounted for [3].

Recently, however, Gibson and co-workers [4] have suggested that $\theta=7^\circ$ may be a more appropriate relaxation angle, based on analyses of Rutherford back-scattering (RBS) data. Duke and co-workers have also presented analyses of LEED data that indicate that a 7° rotation, while not preferred, is acceptable [5]. Gibson et al. have stated, however, that their data might be consistent with the $\theta=27^\circ$ model, provided one allows for anomalously large surface phonon amplitudes.

With LEED and RBS analyses producing ambiguous interpretations of the data, we thought it might be useful to determine if the measured surface state dispersion curves $E(\bar{k})$ [6], when compared with theoretical predictions, preferred either the $\theta=7^\circ$ model or the $\theta=27^\circ$ model. Previous calculations of $E(\bar{k})$, assuming the $\theta=27^\circ$ model, were in sufficiently good agreement with the data to afford explanations of the principal experimental features [7] (Fig.1). However, we now find qualitatively similar dispersion relations $E(\bar{k})$ for the relevant surface states, for $\theta=0^\circ$, $\theta=7^\circ$, 14° and 27° (Fig. 1). Since the theory is only accurate to several tenths of an eV [8] near the valence band maximum, the theoretical surface state dispersion curves do not provide a means for discriminating with confidence among the relaxation models. The theory does predict that surface states do fall in the fundamental band gap for the 7° model [1] (Fig. 2 [9] and Ref. [10]): 0.1 eV below the conduction band edge and 0.1 eV above the valence band maximum — but these energies are too small in comparison with the several tenths of an eV theoretical uncertainty to be convincing proof of the 27° model over the 7° model.

Hence we conclude that the agreement between photoemission data and the theory does not provide strong evidence for or against either the $\theta=7^\circ$ model or the $\theta=27^\circ$ model. The established $\theta=27^\circ$ model should be retained until more conclusive experimental evidence against it is presented.

Finally, as we have been completing this manuscript, we have received a preprint from Mailhiot, Duke and Chang [11], who have independently been studying this problem using the same Hamiltonian and comparably accurate theoretical techniques. They have found similar results; however they interpret their results as providing stronger support for the $\theta=27^\circ$ model.

Acknowledgments — We are grateful to the U.S. Army Research Office and the Office of Naval Research for their support (Contract Nos. DAAG-29-83-K-0122 and N00014-82-K-0447).

1982 PACS Number: 68.20.+t; 73.20.-r

0749-6036/85/010087+03 $02.00/0

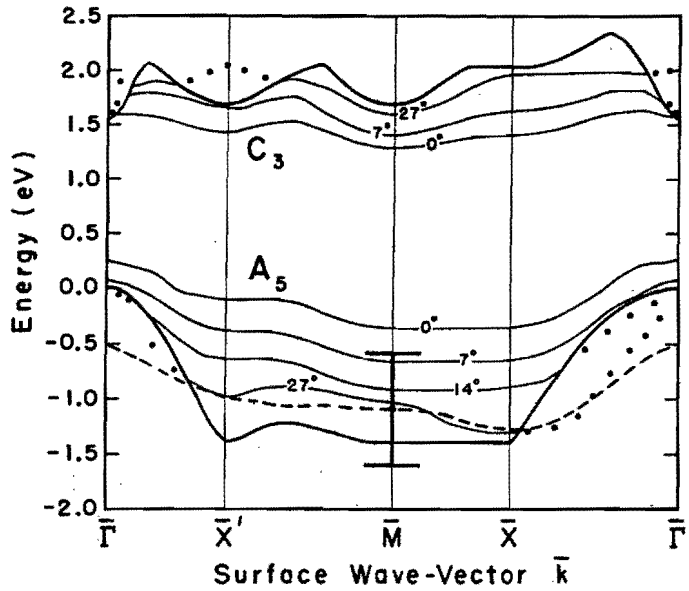


Fig. 1. Calculated surface state dispersion curves in the gap (solid lines) and resonant with the bulk bands (dotted lines), assuming a surface bond rotation angle of $\theta=0^{\circ}$, 7°, 14°, and 27° for the dangling bond (110) surface states C_3 and A_5 of GaAs. Heavy solid lines represent the bulk band edges; the dashed line denotes the data of A. Huijser, J. van Laar and T. L. van Rooy, Phys. Lett. 65A, 337 (1978) and G. P. Williams, R. J. Smith and G. J. Lapeyre, J. Vac. Sci. Technol. 15, 1249 (1978). The C_3 state is not shown for $\theta=14^{\circ}$, because this state lies too close to the $\theta=7^{\circ}$ and 27° states. The 27° results are the same as those of Ref. [7]. The absolute uncertainty in the theoretical predictions is shown by the error bar at $\bar{M}$ centered on the data.

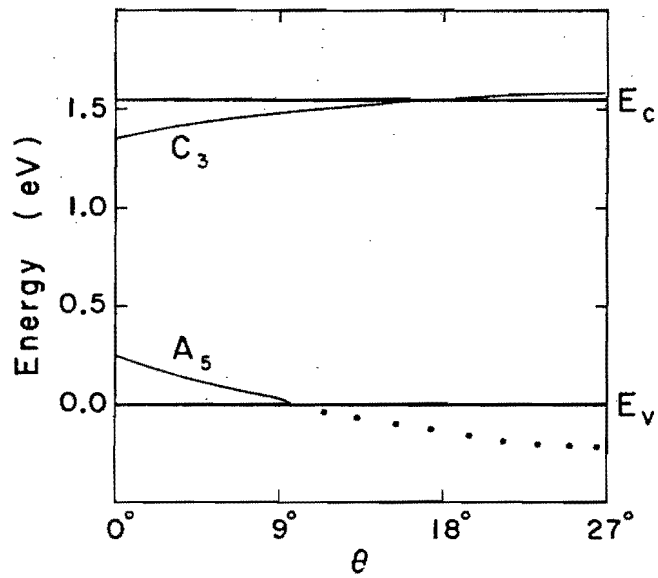


Fig. 2. Calculated energies of the dangling bond (110) surface states of GaAs (solid lines for bound states, dotted lines for resonances) at $\bar{\Gamma}$ [9], as functions of the surface bond rotation angle. The results for the C_3 state are from Ref. [10]. E_v and E_c are the valence and conduction band edges, denoted by heavy solid lines.

REFERENCES

[1] S. Y. Tong, A. R. Lubinsky, B. J. Mrstik and M. A. van Hove, Phys. Rev. B17, 3303 (1978).

[2] A. Kahn, E. So, P. Mark and C. B. Duke, J. Vac. Sci. Technol. 15, 580 (1978); C. B. Duke, R. J. Meyer and P. Mark, J. Vac. Sci. Technol. 17, 971 (1980).

[3] For a sampling of previous theories of the GaAs (110) surface states, see C. Calandra, F. Manghi and C. M. Bertoni, J. Phys. C10, 1911 (1977); J. D. Joannopoulos and M. L. Cohen, Phys. Rev. B10, 5075 (1974); E. J. Mele and J. D. Joannopoulos, Phys. Rev. B17, 1816 (1978); D. J. Chadi, J. Vac. Sci. Technol. 15, 631, 1244 (1978), Phys. Rev. B18, 1800 (1978); J. A. Knapp, D. E. Eastman, K. C. Pandey and F. Patella, J. Vac. Sci. Technol. 15, 1252 (1978); J. R. Chelikowsky and M. L. Cohen, Phys. Rev. B20, 4150 (1979); A. Zunger, Phys. Rev. B22, 959 (1980); A. Mazur, J. Pollmann and M. Schmeits, Solid State Commun. 42, 37 (1982).

[4] W. M. Gibson and H. J. Gossman, J. Vac. Sci. Technol. B2, 343 (1984).

[5] C. B. Duke, S. L. Richardson, A. Paton and A. Kahn, Surf. Science 128, L135 (1983).

[6] J. van Laar and J. J. Scheer, Surface Sci 8, 342 (1967); J. van Laar and A. Huijser J. Vac. Sci. Technol. 13, 769 (1976); V. Gudat and D. E. Eastman, J. Vac. Sci Technol. 13, 831 (1976); W. E. Spicer, I Lindau, P. E. Gregory, C. M. Garner, P Pianetta and P. Chye, J. Vac. Sci. Technol 13, 780 (1976).

[7] R. P. Beres, R. E. Allen and J. D. Dow Solid State Commun. 45, 13 (1983).

[8] The theoretical uncertainty can be ±0.5 eV

[9] To avoid singular matrices, th calculations for $\bar{\Gamma}$ are actually performe at $\bar{k}=0.02\bar{X}$. For $\theta \geq 21^{o}$, the smaller- results for A_5 are extrapolated. Thes results are checked by considering severa different directions of $\bar{k}$, near $\bar{\Gamma}$.

[10] R. E. Allen, H. P. Hjalmarson and J. D. Dc Surf. Sci. 110, L625 (1981).

[11] C. Mailhiot, C. B. Duke and Y. C. Chang "The atomic geometries of ZnSe (110) an GaAs (110): determination by photoemissic spectroscopy," to be published.

RELATIVISTIC EMPIRICAL TIGHT-BINDING THEORY OF THE ENERGY BANDS OF GeTe, SnTe, PbTe, PbSe, PbS, AND THEIR ALLOYS

Craig S. Lent[a], Marshall A. Bowen[b], John D. Dow, and Robert S. Allgaier[c]
Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556 U.S.A.

and

Otto F. Sankey
Department of Physics, Arizona State University, Tempe, Arizona 85287 U.S.A.

and

Eliza S. Ho
Department of Physics, University of Illinois at Urbana-Champaign
Urbana, Illinois 61801 U.S.A.

(Received 29 July 1986)

The orthogonalized plane wave band structures of GeTe, SnTe, PbTe, PbSe, and PbS are fit with a nearest-neighbor, 18-orbital sp^3d^5, relativistic tight-binding model that exhibits chemical trends. The band gaps of $Pb_{1-x}Sn_xTe$, $Sn_{1-y}Ge_yTe$, and $Ge_{1-z}Pb_zTe$ alloys are predicted as functions of compositions x, y, and z. Bowing of the gap is expected to be substantial for $Ge_{1-z}Pb_zTe$, and either $Sn_{1-y}Ge_yTe$ or $Ge_{1-z}Pb_zTe$ should exhibit a Dimmock reversal.

1. Introduction

The rocksalt-structure IV-VI semiconductor compounds, such as PbTe, SnTe, GeTe, PbSe, and PbS all have small band gaps, high dielectric constants, interesting defect levels, and a variety of very unusual thermodynamic, vibrational, electronic, and infrared properties [1]. Exploitation of these properties for the fabrication of technologically important opto-electronic devices has been partially impeded by an incomplete understanding of the intrinsic and extrinsic electronic states of these materials. The IV-VI's have attracted relatively little theoretical attention, however, because their electronic band structures are complicated, having large relativistic splittings. At first glance, it would appear that the electronic states of bulk defects or surfaces of these materials can be understood only if one executes a very tedious, relativistic theory.

In this paper, we show that the apparently complicated energy bands of the IV-VI compounds can be parameterized by a simple nearest-neighbor tight-binding model Hamiltonian. The parameters of this model exhibit chemical trends and can be used to predict the electronic structures of alloys such as $Pb_{1-x}Sn_xTe$. Moreover, theories of defect energy levels and surface states in IV-VI's can be constructed using this simple Hamiltonian, as we shall demonstrate in subsequent work.

2. Tight-binding theory

The relativistic Hamiltonian that produces the energy band structures has the form [2]

$$H = (p^2/2m) + V + H_{so} + \hbar^2\nabla^2 V/8m^2c^2 - p^4/8m^3c^2 \quad (1)$$

where V is the crystal potential, the spin orbit interaction is

$$H_{so} = \hbar\vec{\sigma}\cdot(\nabla V\times\vec{p})/4m^2c^2,$$

and the remaining terms are the Darwin terms and the relativistic mass correction term [3].

Employing the ideas of Slater and Koster [4], Harrison [5], Chadi [6], and Vogl et al. [7], we construct the nearest-neighbor tight-binding Hamiltonian:

$$H_0 = \Sigma_{\vec{R},\sigma,i}\,[|a,i,\sigma,\vec{R}\rangle E_{i,a}\langle a,i,\sigma,\vec{R}| + |c,i,\sigma,\vec{R}+\vec{d}\rangle E_{i,c}\langle c,i,\sigma,\vec{R}+\vec{d}|] + \Sigma_{\vec{R},\vec{R}',\sigma,i,j}\,[|a,i,\sigma,\vec{R}\rangle V_{i,j}\langle c,j,\sigma,\vec{R}'+\vec{d}| + \text{h.c.}] + H_{so}, \quad (2)$$

(a) Permanent address: Department of Electrical and Computer Engineering, University of Notre Dame, Notre Dame, Indiana 46556.
(b) Permanent address: Department of Computer Science, Western Illinois University, Macomb, Illinois 61455.
(c) Present address: Theodore Associates, Inc., 10510 Streamview Court, Potomac, Maryland 20854.

PACS Number: 71.25.Tn

0749-6036/86/050491+09 $02.00/0

TABLE I. Nearest-neighbor tight-binding parameters of GeTe, SnTe, PbTe, PbSe, and PbS, as fit to the band structure of Herman et al. [9], in eV. The column labelled GeTe* refers to Dimmock-reversed GeTe with the valence (conduction) band extremum at L_6^- (L_6^+). $V_{d,d\pi}$, $V_{d,s}$, and $V_{s,d}$ are taken to be zero.

	GeTe	GeTe*	SnTe	PbTe	PbSe	PbS
$E_{s,c}$	-7.847	-7.992	-6.578	-7.612	-7.010	-6.546
$E_{s,a}$	-10.974	-10.855	-12.067	-11.002	-13.742	-13.827
$E_{p,c}$	1.454	1.657	1.659	3.195	4.201	3.486
$E_{p,a}$	0.444	0.250	-0.167	-0.237	-1.478	-1.153
$E_{d,c}$	9.08	9.08	8.38	7.73	8.72	9.27
$E_{d,a}$	25.85	26.75	7.73	7.73	11.95	10.38
λ_c	0.505	0.577	0.592	1.500	1.693	1.559
λ_a	0.447	0.351	0.564	0.428	0.121	-0.211
$V_{s,s}$	-0.617	-0.631	-0.510	-0.474	-0.402	-0.364
$V_{s,p}$	0.877	0.788	0.949	0.705	0.929	0.936
$V_{p,s}$	0.790	0.876	-0.198	0.633	0.159	0.186
$V_{p,p}$	2.189	2.I81	2.218	2.066	1.920	2.073
$V_{p,p\pi}$	-0.478	-0.498	-0.446	-0.430	-0.356	-0.281
$V_{p,d}$	-1.14	-1.65	-1.11	-1.29	-1.590	-1.142
$V_{p,d\pi}$	1.56	1.78	0.624	0.835	1.45	1.16
$V_{d,p}$	-1.55	-1.50	-1.67	-1.59	-1.09	-1.54
$V_{d,p\pi}$	0.976	0.742	0.766	0.531	0.0497	0.517
$V_{d,d}$	-3.79	-3.87	-1.72	-1.35	-1.90	-1.67
$V_{d,d\delta}$	0.887	0.892	0.618	0.668	0.692	0.659

Table II. Experimental values of the fundamental gap for GeTe, SnTe, PbTe, PbSe, and PbS used in fitting the tight-binding parameters of Table I (in eV).

	GeTe	SnTe	PbTe	PbSe	PbS
E_{gap}	0.2[a]	0.3[b]	0.186[c]	0.165[d]	0.286[d]

[a] L. Esaki, J. Phys. Soc. Japan, 1966, 21, 589 [Kyoto Conference Supplement], measurements at 4.2°K.
[b] Ref. [10], measurements at 4.2°K.
[c] Ref. [10], measurements at 12°K.
[d] D. L. Mitchell, E. D. Palik, and J. N. Zemel, Proc. Seventh Int. Conf. Phys. Semicond., 1964, p. 325 (1964), measurements at 4.2°K.

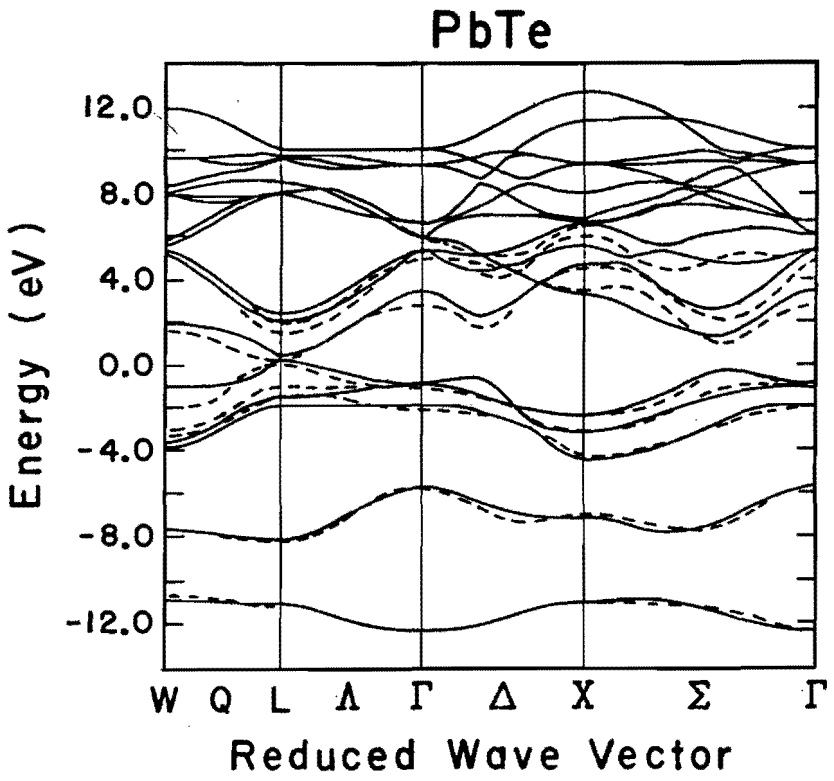


Fig. 1. The energy band structure in eV of PbTe, published by Herman et al. [9] (dashed) in comparison with the present work (solid). Note that the zero of energy is the valence band maximum and that the fundamental band gap is at L.

where h.c. means Hermitian conjugate, $\vec{R}$ are the rock-salt lattice positions of the anion, i and j are the basis orbitals for the cation and anion respectively, σ is the spin (up or down), a and c refer to the anion and cation respectively, and $\vec{d}$ is the position of the cation relative to the anion in the $\vec{R}$-th cell; $\vec{d} = (a_L/2)(1,0,0)$. The spin-orbit Hamiltonian is

$$H_{so} = \Sigma_{\vec{R},\sigma,\sigma',i} [|c,i,\sigma,\vec{R}\rangle \lambda_c \vec{L}_c \cdot \vec{\sigma}_c \langle c,i,\sigma',\vec{R}|] + \Sigma_{\vec{R},\sigma,\sigma',j} [|a,j,\sigma,\vec{R}\rangle \lambda_a \vec{L}_a \cdot \vec{\sigma}_a \langle a,j,\sigma',\vec{R}|]. \quad (3)$$

We use nine orbitals per atom in our basis, each with up and down spin: s, p_x, p_y, p_z, $d_{x^2-y^2}$, $d_{3z^2-r^2}$, d_{xy}, d_{yz}, d_{zx}. Because of the importance of the d bands near the bottom of the conduction band at the X point we found it necessary to include all five d bands in the model. This approach is to be preferred over that of Robertson [8], which included only two of the five d orbitals. We did neglect (i) the somewhat smaller couplings $V_{s,d\sigma}$ between the s states and the d states and (ii) $V_{d,d\pi}$, the π-type bonding between d states.

The resulting 36×36 Hamiltonian matrix is given in Appendix A.

3. Determination of the empirical Hamiltonian matrix elements

The parameters of this model are listed in Table I. They were obtained by fitting the eigenvalues of the matrix to the energy bands published by Herman et al. [9] (See Fig. 1). Analytic expressions for the eigenvalues at high symmetry points were used to make an initial guess for the parameters. Then a least-squares fit of the parameters to the calculated energy bands was performed. The symmetry of the states on either side of the fundamental gap was also included in the fitting procedure. This is necessary to assure the Dimmock reversal [10] in the ordering of bands that occurs in $Pb_{1-x}Sn_xTe$ between PbTe (with a conduction band minimum at L_6^- and valence band maximum at L_6^+) and SnTe (with the opposite ordering). The energy bands were fit to the values obtained by Herman et al. for wavevectors at the Γ, X, and L points of the Brillouin zone; but Herman's conduction band energies at L were all shifted by the same small amount in order to guarantee that the fundamental band gap agreed with experiment. The resulting band structures are displayed in Figs. 1-5. The fit of the band structure of GeTe

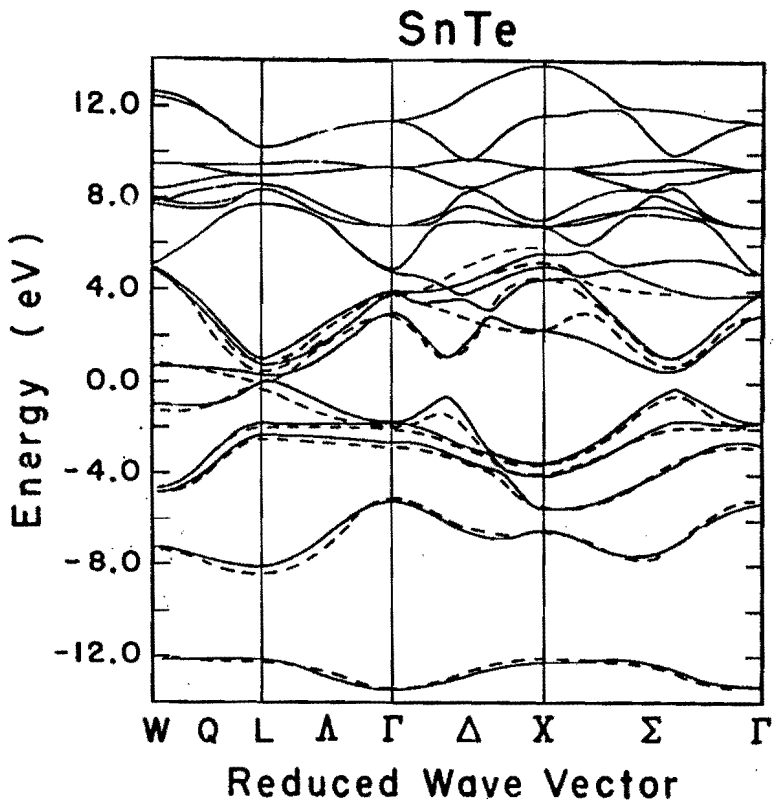


Fig. 2. The calculated energy bands of SnTe.

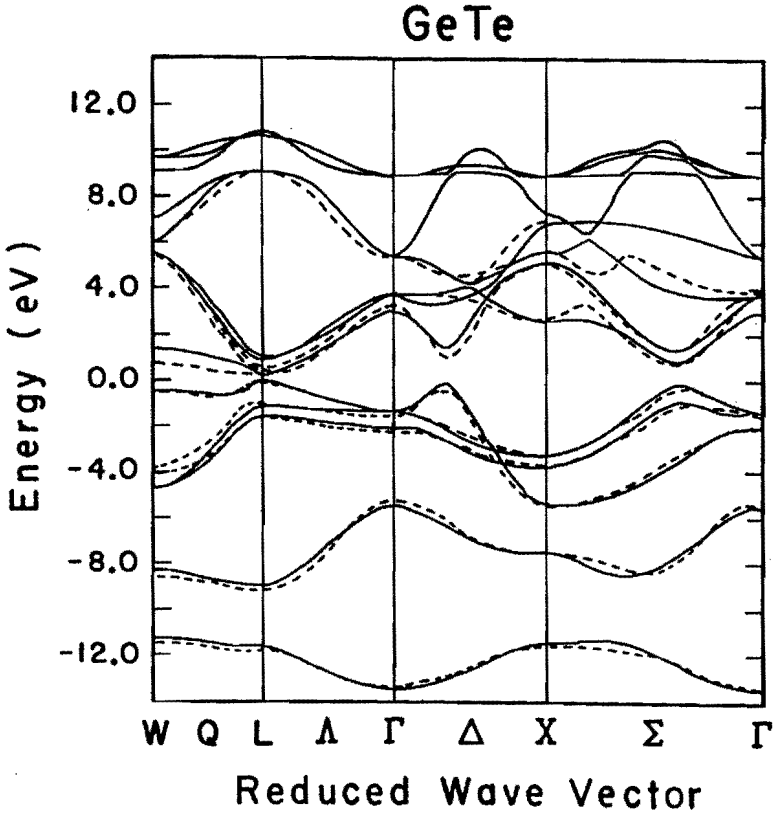


Fig. 3. The calculated energy bands of cubic (rocksalt) GeTe.

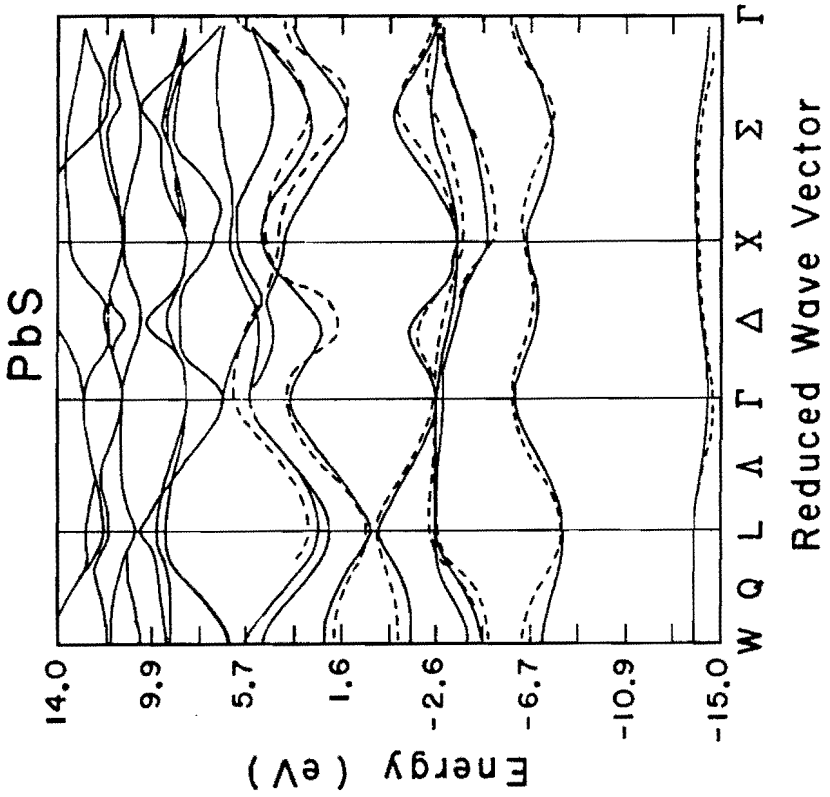


Fig. 5. The calculated energy bands of PbS.

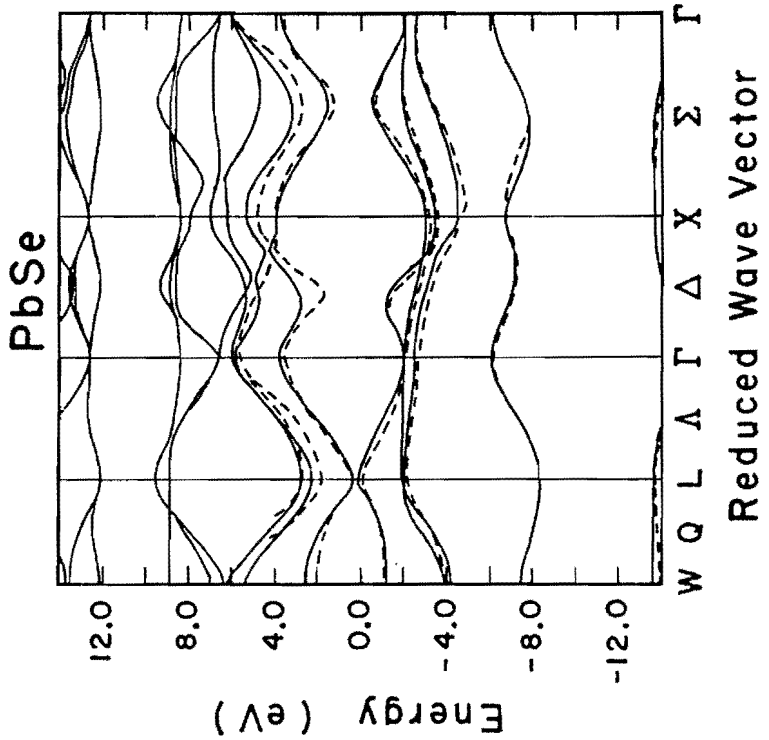


Fig. 4. The calculated energy bands of PbSe.

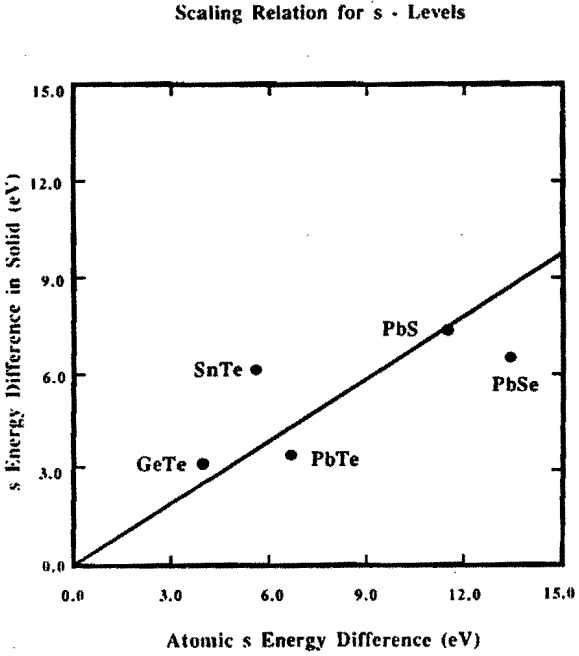


Fig. 6. The s-orbital energy differences in the solid, (Table I) versus the s-orbital energy differences in the atom [7].

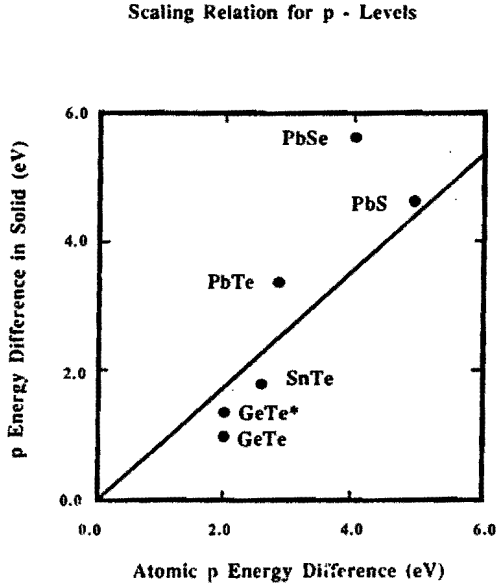


Fig. 7. The p-orbital energy differences in the solid, (Table I) versus the p-orbital energy differences in the atom [7].

assumes a PbTe-like ordering of the conduction and valence bands: L_6^- above L_6^+. The possibility exists, however, that GeTe's band structure is Dimmock-reversed, as SnTe's is, with L_6^+ being the conduction band minimum. We denote Dimmock-reversed GeTe by GeTe*, and obtain for it the slightly different matrix elements listed in Table I.

For the parameters of the model, the differences in the diagonal matrix elements $E_{s,c}-E_{s,a}$ and $E_{p,c}-E_{p,a}$, are approximately proportional to the corresponding differences of atomic energies. The Vogl constant of proportionality β [7] is about 0.65 for the s-state and 0.9 for the p-state. (See Figs. 6 and 7.) These proportionalities or scaling rules for the matrix elements of the empirical Hamiltonian allow the theory to make sensible predictions of chemical trends for intrinsic and extrinsic electronic states of different IV-VI semiconductors.

4. Application to Alloys

In this section we apply the theory to $Pb_{1-x}Sn_xTe$, $Sn_{1-y}Ge_yTe$, and $Ge_{1-z}Pb_zTe$ alloys and compute the alloy band gaps as functions of the compositions x, y, and z, using the virtual crystal approximation. These materials are substitutional alloys miscible for all compositions.

$Pb_{1-x}Sn_xTe$ is an interesting alloy because the band gap of SnTe is "inverted" in comparison with the gap of PbTe: the fundamental band gap occurs with the conduction band minimum being the L_6^+ point of the band structure, rather than at the the L_6^- point. This phenomenon was elucidated by Dimmock et al. [10] several years ago, who pointed out that a level-crossing phenomenon occurs with increasing x as the band-gap of $Pb_{1-x}Sn_xTe$ decreases and attempts to become negative. We calculate that the gap vanishes at $x \approx 0.35$, in good agreement with the experimental value.

As a function of alloy composition, this Dimmock reversal in $Pb_{1-x}Sn_xTe$ must undo itself in either $Sn_{1-y}Ge_yTe$ or $Ge_{1-z}Pb_zTe$. We predict that the second Dimmock reversal must occur either near $y \approx 0.6$ in $Sn_{1-y}Ge_yTe$ or near $z \approx 0.3$ in $Ge_{1-z}Pb_zTe$. In the former case, GeTe must have the same ordering of L_6 bands as PbTe, whereas in the latter case, GeTe has the SnTe-like GeTe* electronic structure. (See Fig. 8.) The calculations also indicate that one should expect considerable bowing in the fundamental band gap versus alloy composition for $Ge_{1-z}Pb_zTe$, in contrast to the linear x-dependence of the gap for $Pb_{1-x}Sn_xTe$. This striking prediction of the calculations is in qualitative agreement with the measurements of Nikolic [11,12].

We compute the fundamental band gaps of alloys such as $Pb_{1-x}Sn_xTe$ by diagonalizing the virtual-crystal [13] empirical tight-binding Hamiltonian. The covalent radii of Pb and Sn differ by so little (≈ 4%) and all of the Hamiltonian matrix elements of PbTe and SnTe are sufficiently similar that a virtual crystal

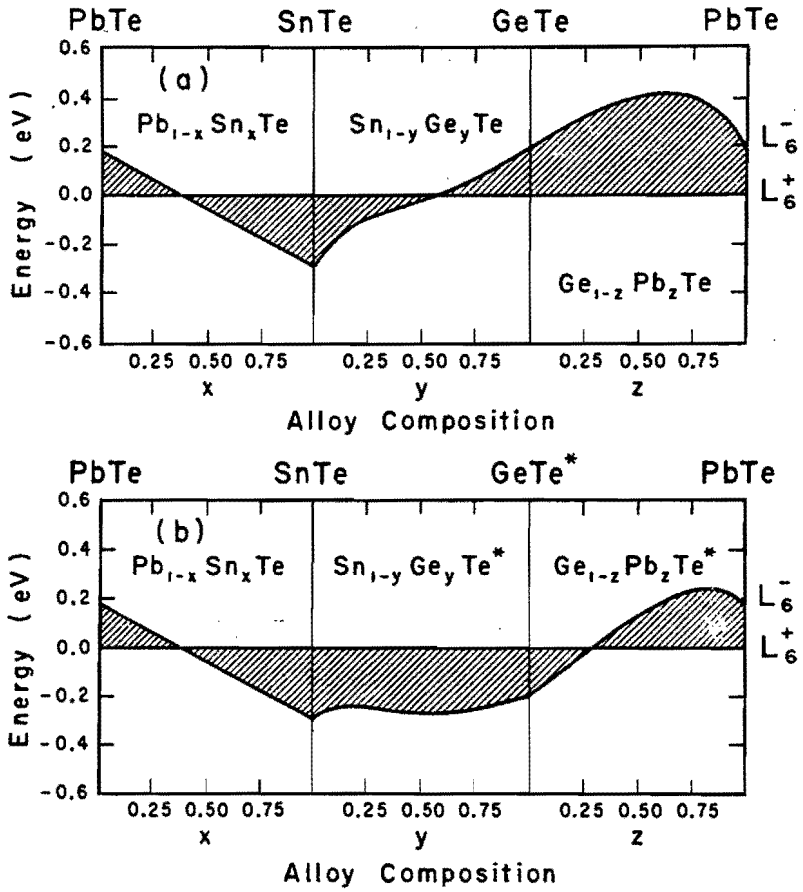


Fig. 8. The calculated band gaps and L_6^+ and L_6^- band edges of $Pb_{1-x}Sn_xTe$, $Sn_{1-y}Ge_yTe$, and $Ge_{1-z}Pb_zTe$ versus alloy compositions x, y, and z. The zero of energy is the L_6^+ band extremum. The hatched area is the gap. (a) For ordinary GeTe; (b) for Dimmock-reversed GeTe* (see text).

approximation to the matrix elements of $Pb_{1-x}Sn_xTe$ is appropriate for states near the fundamental band gap [13]: the diagonal Hamiltonian matrix elements of $Pb_{1-x}Sn_xTe$ are (1-x) times the PbTe elements plus x times the SnTe matrix elements [14]. The off-diagonal matrix elements, multiplied by the square of the lattice constant [5], are similarly averaged, using Vegard's Law for the lattice constant. The band edges of Fig. 8 were eigenvalues obtained by diagonalizing this Hamiltonian for a wavevector at the L-point of the Brillouin zone.

5. Conclusion

We conclude that the present tight-binding parameterization of the IV-VI energy bands is adequate for reproducing chemical trends, including the Dimmock band reversal phenomenon in $Pb_{1-x}Sn_xTe$. Therefore, it should provide a satisfactory starting point for general theories of localized electronic states in these very small band-gap materials, such as "deep traps" [15] or surface states. Subsequent work will use this Hamiltonian to study a wide variety of problems involving localized electronic states in IV-VI semiconductors.

Acknowledgments -- The authors thank the Office of Naval Research and the Army Research Office for the support of this work (Contract Nos. N00014-84-K-0352 and ARO-DAAG29-81-K-0068). One of us (E. S. H.) was supported in part by the

Department of Energy, Division of Materials Science (Contract No. DE-AC02-76ER01198) and another (R. S. A.) was supported in part by the Independent Research Fund of the Naval Surface Weapons Center. We thank R. Hong for carefully proof-reading the manuscript and checking the tight-binding parameters. We are especially grateful to G. B. Wright, whose enthusiastic support and encouragement led to the initiation of this work.

APPENDIX A: The 36×36 Hamiltonian matrix

The basis set for the 36×36 Hamiltonian is: $|s,c,\uparrow\rangle$, $|s,c,\downarrow\rangle$, $|s,a,\uparrow\rangle$, $|s,a,\downarrow\rangle$, $|p_x,c,\uparrow\rangle$, $|p_y,c,\uparrow\rangle$, $|p_z,c,\uparrow\rangle$, $|p_x,c,\downarrow\rangle$, $|p_y,c,\downarrow\rangle$, $|p_z,c,\downarrow\rangle$, $|p_x,a,\uparrow\rangle$, $|p_y,a,\uparrow\rangle$, $|p_z,a,\uparrow\rangle$, $|p_x,a,\downarrow\rangle$, $|p_y,a,\downarrow\rangle$, $|p_z,a,\downarrow\rangle$, $|d_1,c,\uparrow\rangle$, $|d_2,c,\uparrow\rangle$, $|d_3,c,\uparrow\rangle$, $|d_4,c,\uparrow\rangle$, $|d_5,c,\uparrow\rangle$, $|d_1,c,\downarrow\rangle$, $|d_2,c,\downarrow\rangle$, $|d_3,c,\downarrow\rangle$, $|d_4,c,\downarrow\rangle$, $|d_5,c,\downarrow\rangle$, $|d_1,a,\uparrow\rangle$, $|d_2,a,\uparrow\rangle$, $|d_3,a,\uparrow\rangle$, $|d_4,a,\uparrow\rangle$, $|d_5,a,\uparrow\rangle$, $|d_1,a,\downarrow\rangle$, $|d_2,a,\downarrow\rangle$, $|d_3,a,\downarrow\rangle$, $|d_4,a,\downarrow\rangle$, and $|d_5,a,\downarrow\rangle$,

where we have $d_1 = d_{x^2-y^2}$, $d_2 = d_{3z^2-r^2}$, $d_3 = d_{xy}$, $d_4 = d_{yz}$, and $d_5 = d_{zx}$.

The Hamiltonian can be written in block form as follows (only the lower triangular part of the Hamiltonian is given since it is Hermitian):

$$\begin{bmatrix} H_{s,s} & & & & \\ H_{pc,s} & H_{pc,pc} & & & \\ H_{pa,s} & H_{pa,pc} & H_{pa,pa} & & \\ 0 & 0 & H_{dc,pa} & H_{dc,dc} & \\ 0 & H_{da,pc} & 0 & H_{da,dc} & H_{da,da} \end{bmatrix}$$

$H_{s,s}$ is a Hermitian 4×4 matrix which connects s-states to s-states:

$$H_{s,s} = \begin{bmatrix} E_{s,c} & & & \\ 0 & E_{s,c} & & \\ g_0V_{ss} & 0 & E_{s,a} & \\ 0 & g_0V_{s,s} & 0 & E_{s,a} \end{bmatrix}$$

and $H_{pc,s}$ can be written as:

$$H_{pc,s} = \begin{bmatrix} 0 & H_1 \\ 0 & H_2 \end{bmatrix}$$

where we have

$$H_1 = \begin{bmatrix} -2g_1V_{p,s} & 0 \\ -2g_2V_{p,s} & 0 \\ -2g_3V_{p,s} & 0 \end{bmatrix}$$

and

$$H_2 = \begin{bmatrix} 0 & -2g_1V_{p,s} \\ 0 & -2g_2V_{p,s} \\ 0 & -2g_3V_{p,s} \end{bmatrix}$$

$H_{pa,s}$ can be written as:

$$H_{pa,s} = \begin{bmatrix} H_3 & 0 \\ H_4 & 0 \end{bmatrix}$$

where we have

$$H_3 = \begin{bmatrix} -2g_1V_{s,p} & 0 \\ -2g_2V_{s,p} & 0 \\ -2g_3V_{s,p} & 0 \end{bmatrix}$$

and

$$H_4 = \begin{bmatrix} 0 & -2g_1V_{s,p} \\ 0 & -2g_2V_{s,p} \\ 0 & -2g_3V_{s,p} \end{bmatrix}.$$

$H_{pc,pc}$ is a 6×6 Hermitian matrix with all the diagonal elements equal to $E_{p,c}$. Its other non-zero matrix elements in the lower triangular region are:

$$\langle p_y,c,\uparrow| H |p_x,c,\uparrow\rangle = i\,\lambda_c/2,$$
$$\langle p_z,c,\downarrow| H |p_x,c,\uparrow\rangle = \lambda_c/2,$$
$$\langle p_z,c,\downarrow| H |p_y,c,\uparrow\rangle = i\,\lambda_c/2,$$
$$\langle p_x,c,\downarrow| H |p_z,c,\uparrow\rangle = -\lambda_c/2,$$
$$\langle p_y,c,\downarrow| H |p_z,c,\uparrow\rangle = -i\lambda_c/2,$$

and

$$\langle p_y,c,\downarrow| H |p_x,c,\downarrow\rangle = -i\lambda_c/2.$$

$H_{pa,pa}$ is a 6×6 Hermitian matrix of the same form as $H_{pc,pc}$, but with the diagonal elements equal to $E_{p,a}$ and the other non-zero elements as above with λ_c replaced with λ_a.

$H_{pa,pc}$ is a diagonal 6×6 matrix with,

$$\langle p_x,a,\uparrow| H |p_x,c,\uparrow\rangle = V_{x,x}$$
$$\langle p_y,a,\uparrow| H |p_y,c,\uparrow\rangle = V_{y,y}$$
$$\langle p_z,a,\uparrow| H |p_z,c,\uparrow\rangle = V_{z,z}$$
$$\langle p_x,a,\downarrow| H |p_x,c,\downarrow\rangle = V_{x,x}$$
$$\langle p_y,a,\downarrow| H |p_y,c,\downarrow\rangle = V_{y,y}$$
$$\langle p_z,a,\downarrow| H |p_z,c,\downarrow\rangle = V_{z,z}$$

and

$$V_{x,x} = 2g_4V_{p,p} + 2(g_5+g_6)\,V_{p,p\pi}$$

$$V_{y,y} = 2g_5V_{p,p} + 2(g_4+g_6)\,V_{p,p\pi}$$

$$V_{z,z} = 2g_6V_{p,p} + 2(g_4+g_5)\,V_{p,p\pi}$$

$H_{da,pc}$ can be written in block form as:

$$H_{da,pc} = \begin{bmatrix} H_5 & 0 \\ 0 & H_5 \end{bmatrix}$$

where H_5 is the 5×3 matrix:

$$\begin{bmatrix} -\sqrt{3}g_1V_{p,d} & \sqrt{3}g_2V_{p,d} & 0 \\ g_1V_{p,d} & g_2V_{p,d} & -2g_3V_{p,d} \\ -2g_2V_{p,d\pi} & -2g_1V_{p,d\pi} & 0 \\ 0 & -2g_3V_{p,d\pi} & -2g_2V_{p,d\pi} \\ -2g_3V_{p,d\pi} & 0 & -2g_1V_{p,d\pi} \end{bmatrix}$$

$H_{dc,pa}$ is of the same form, but with $V_{p,d}$ and $V_{p,d\pi}$ replaced by $V_{d,p}$ and $V_{d,p\pi}$.

$H_{da,dc}$ is a 10×10 matrix with only four non-zero off-diagonal elements. The diagonal elements are:

$$\langle d_1,a,\uparrow|\ H\ |d_1,c,\uparrow\rangle = \langle d_1,a,\downarrow|\ H\ |d_1,c,\downarrow\rangle$$
$$= 3/2\ (g_4+g_5)V_{d,d} + (2g_6+g_4/2+g_5/2)V_{d,d\delta}$$
$$\langle d_2,a,\uparrow|\ H\ |d_2,c,\uparrow\rangle = \langle d_2,a,\downarrow|\ H\ |d_2,c,\downarrow\rangle$$
$$= 3/2\ (g_4+g_5)\ V_{d,d\delta} + (2g_6+g_4/2+g_5/2)V_{d,d}$$
$$\langle d_3,a,\uparrow|\ H\ |d_3,c,\uparrow\rangle = \langle d_3,a,\downarrow|\ H\ |d_3,c,\downarrow\rangle$$
$$= 2\ (g_4+g_5)\ V_{d,d\pi} + 2\ g_6\ V_{d,d\delta}$$
$$\langle d_4,a,\uparrow|\ H\ |d_4,c,\uparrow\rangle = \langle d_4,a,\downarrow|\ H\ |d_4,c,\downarrow\rangle$$
$$= 2\ (g_5+g_6)\ V_{d,d\pi} + 2\ g_4\ V_{d,d\delta}$$
$$\langle d_5,a,\uparrow|\ H\ |d_5,c,\uparrow\rangle = \langle d_5,a,\downarrow|\ H\ |d_5,c,\downarrow\rangle$$
$$= 2\ (g_4+g_6)\ V_{d,d\pi} + 2\ g_5\ V_{d,d\delta}.$$

The non-zero off-diagonal elements are all equal:

$$\langle d_1,a,\uparrow|\ H\ |d_2,c,\uparrow\rangle = \langle d_2,a,\uparrow|\ H\ |d_1,c,\uparrow\rangle$$
$$= \langle d_1,a,\downarrow|\ H\ |d_2,c,\downarrow\rangle = \langle d_2,a,\downarrow|\ H\ |d_1,c,\downarrow\rangle$$
$$= (\sqrt{3})/2\ (g_5-g_4)\ (V_{dd} - V_{d,d\delta})$$

where we have

$$g_0(\vec{k}) = 2[\cos(k_xa_L/2)+\cos(k_ya_L/2)+\cos(k_za_L/2)],$$
$$g_1(\vec{k}) = i\ \sin(k_xa_L/2),$$
$$g_2(\vec{k}) = i\ \sin(k_ya_L/2),$$
$$g_3(\vec{k}) = i\ \sin(k_za_L/2),$$
$$g_4(\vec{k}) = \cos(k_xa_L/2),$$
$$g_5(\vec{k}) = \cos(k_ya_L/2),$$

and

$$g_6(\vec{k}) = \cos(k_za_L/2).$$

The parameters $V_{s,s}$, $V_{s,p}$ and $V_{p,d}$ correspond to the integrals $(ss\sigma)_1$, $(sp\sigma)_1$, $(pd\sigma)_1$ in Ref. [4].

$H_{d,c}$ and $H_{d,a}$ are both 10×10 diagonal matrices whose elements are $E_{d,c}$ and $E_{d,a}$ respectively.

REFERENCES

[1] G. Nimtz and B. Schlicht in Narrow Gap Semiconductors, (Springer Tracts in Modern Physics Vol. 98, Springer-Verlag, Berlin, 1983, ed. G. Höhler), p. 1.

[2] N. J. Parada, Phys. Rev. B3, 2042 (1971).

[3] H. A. Bethe and E. E. Salpeter, Quantum Mechanics of One- and Two-electron Atoms, (Springer-Verlag, Berlin, 1957), p. 56.

[4] J. C. Slater and G. F. Koster, Phys. Rev. 94, 1498 (1954).

[5] W. A. Harrison, Phys. Rev. B8, 4487 (1973); W. A. Harrison, Electronic Structure and the Properties of Solids, (W. H. Freeman and Company, San Francisco, 1980).

[6] D. J. Chadi, Phys. Rev. B16, 790 (1977).

[7] P. Vogl, H. P. Hjalmarson and J. D. Dow, J. Phys. Chem. Solids 44, 365 (1983).

[8] J. Robertson, Phys. Rev. B28, 4671 (1983).

[9] F. Herman, R. L. Kortum, I. B. Ortenburger and J. P. Van Dyke, J. Physique, 29, Suppl. C4, 62 (1968).

[10] J. O. Dimmock, I. Melngailis and A. J. Strauss, Phys. Rev. Letters, 16, 1193 (1966).

[11] The theory cannot confidently predict the bowing with quantitative accuracy. Rather it suggests that strong bowing should occur for $Ge_{1-z}Pb_zTe$ with z near unity.

[12] P. M. Nikolic, J. Phys. D 2, 383 (1969).

[13] Y. Onodera and Y. Toyozawa, J. Phys. Soc. Jpn. 24, 341 (1968). The range of validity of the virtual crystal approximation is a subject of current controversy. For summaries of the differing viewpoints, see L. C. Davis, Phys. Rev. B28, 6961 (1983) and references therein, and S. Lee, Ph.D. thesis, Physics Department, University of Illinois, 1986, and references therein.

[14] We fit the band gap of PbTe to experimental results at 12°K, but that of SnTe to results at 4.2K°, as given by Dimmock [10]. However, the changes in the gaps between the two temperatures amounts to only 1 or 2 meV, for both PbTe and SnTe. See the empirical relationship on page 42 of Ref. [1].

[15] C. S. Lent, M. A. Bowen, R. S. Allgaier, J. D. Dow, O. F. Sankey, and E. S. Ho, "Impurity levels in PbTe and $Pb_{1-x}Sn_xTe$," to be published.

Solid State Communications, Vol.51,No.9, pp.705-708, 1984
Printed in Great Britain.

0038-1098/84 $3.00 + .00
Pergamon Press Ltd.

THEORY OF ELECTRON SPIN RESONANCE MEASUREMENTS OF CHALCOGEN PAIRS IN Si

Otto F. Sankey
Department of Physics, Arizona State University, Tempe, Arizona 85287 U.S.A.

and

John D. Dow
Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556 U.S.A.

(Original version received 11 August 1983; revised version received 6 March 1984 by J. Tauc)

Calculations are presented which support the identification of two sulfur-related centers in Si (lying 0.37 and 0.19 eV below the conduction band minima) with $(S,S)^+$ and $(S,S)^0$ nearest-neighbor substitutional S pairs. Explanations in terms of meso-bonding are given of the following facts: (i) Although S is much more electronegative than Si, the $(S,S)^+$ pair level lies at higher energy than the S^+ level by ≈0.2 eV; (ii) The hyperfine interaction for $(S,S)^+$ is considerably smaller than for the isolated S defect; and (iii) The $(S,S)^+$ molecular defect has a hyperfine tensor that is virtually isotropic.

Nearly twenty years ago, Ludwig observed evidence of substitutional nearest-neighbor S pairs in Si, while studying the electron spin resonance and electron nuclear double resonance spectra of isolated S [1]. It is now known that there are five S-related deep levels in the band gap of Si [2-7], two of which are associated with isolated-S (the A and B centers [8]), two of which are thought to be caused by paired-S defects (hereafter denoted the C and D centers [8]), and the fifth Z center [9], which is associated with unknown S-related complexes. The energy levels of the A, B, C, and D centers and their assignments are shown in Fig. 1.

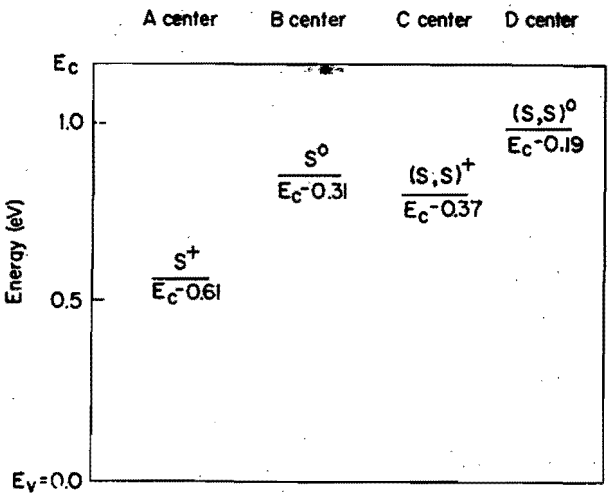


Fig. 1. The observed sulfur-related deep energy levels in the band gap of Si, together with the defect assignments of Brotherton and earlier workers. These assignments are confirmed by the present work.

Brotherton et al. [9] have shown that high temperature annealing leads to preferential formation of (S,S) pairs, whereas isolated-S defects dominate in material annealed at low temperature. However there are several features of the (S,S) pair identification that, at first glance, appear to be troubling: (i) The pair levels (S,S) lie at higher energy than the isolated-S levels, even though S is very electronegative with respect to Si and replacement of a neighboring Si by a more electronegative atom would be expected to pull the isolated-S level down in energy; (ii) S is so different from Si that in isolation it produces a deep level in the gap; yet the second (spectator) S atom appears to be very little different in its effect from the Si it replaces: it alters the isolated-S deep level by only ≈0.2 eV, even though the additional potential of the second S impurity is strongly attractive (corresponding to a charge of two protons) and localized. The small size of this effect is evident when compared with the even larger difference of deep level energies of S^0 and S^+, ≈0.3 eV, which is due to a single delocalized electron. Also the difference in (S,S) and S energy levels, 0.2 eV, is comparable with the difference in the S and Te deep levels (0.2 eV), even though the defect potential difference of S and Te is only ≈-1 eV compared with -6 eV for S and Si [10]; (iii) Although the (S,S) and S defects have similar energy levels, the hyperfine constant (∝ s-orbital charge density at the sulfur nucleus or $|\psi^S_{3s}(0)|^2$) of (S,S) is smaller by a factor of ≈3 than that of S; and (iv) The hyperfine tensor of (S,S) appears to be isotropic, even though (S,S) is supposedly an oriented molecule.

In this paper we show that these puzzling facts can be explained simply: The assignment of the E_c-0.37 eV and E_c-0.19 eV (C and D) levels to (S,S) pairs is correct, but the relevant energy level of the pair has a completely

different character from that of isolated S. The second S atom drives the first isolated-S deep level into the bottom of the band gap or out of the gap into the valence band — and simultaneously pulls a second level down from the conduction band into the gap (Fig. 2). The first (lower [11]) level we term the totally antibonding (ungerade) state; the second (higher) level we call the meso-bonding (gerade) state, because it is a "bonding" linear combination of the "antibonding" deep levels [10] of the isolated-S defects. The antibonding state is lower in energy for the (S,S) defect (contrary to what one normally finds for diatomic moelcules) because each S deep level has an antibonding host-like wavefunction (rather than a wavefunction similar to atomic sulfur's).

The $(S,S)^+$ level in the gap, predicted using the theory described below, is in excellent agreement with the datum (Fig. 2) and lies higher in energy than isolated-S^+, because the observed pair level is derived from the T_2 (p-like) deep resonance of isolated-S (which is pulled down into the gap by the second S) and not from the A_1 (s-like) level of isolated-S, which is driven down toward the valence band by the second S atom (Fig. 2). The level-repulsion obviously occurring in Fig. 2 is responsible for the $(S,S)^+$ deep level lying very close to the S deep level [12].

The measured ESR hyperfine constant ($|A|=38.4\times10^{-4}$ cm^{-1} [1]) of $(S,S)^+$ is much different from that of S^+ ($|A|=104.2\times10^{-4}$cm^{-1} [1]) because the meso-bonding level has a different character from the A_1 level of S, being derived from the p-like T_2 (resonant) level of isolated-S. If the hyperfine constant of the totally antibonding (lower) level could be measured, it would be near that of S^+ -- we predict it to be only ≈28% smaller. The observed $(S,S)^+$ hyperfine constant is small because the meso-bonding state is less localized on the defect atoms. We calculate the s-orbital

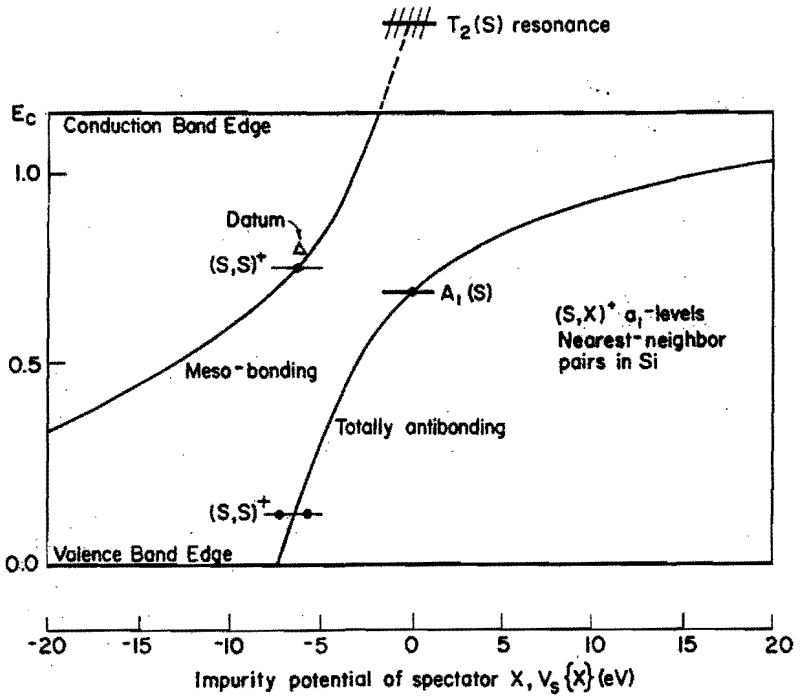


Fig. 2. Predicted energies of deep a_1-symmetric levels associated with $(S,X)^+$ nearest-neighbor substitutional pairs in Si as a function of the defect potential $V_s\{X\}$ of the spectator impurity X. $V_s\{X\}=0$ corresponds to an isolated-S defect, which has a deep level of A_1 (s-like) symmetry in the gap and a T_2 (p-like) deep resonance depicted schematically above the conduction band edge. [Because we are interested primarily in chalcogen pairs, we have assigned one-half of the charge to X and one-half to S in determining the defect potential (See text). Hence the isolated-S A_1 level corresponds to $S^{+1/2}$ and lies between the S^0 and S^+ levels.] Note the excellent agreement between the experimental $(S,S)^+$ level (triangle) and the theory. The electronic occupancies of the one-electron levels of $(S,S)^+$ and S^+ are denoted by circles.

contribution to be 0.33 times as large as that for S^+ — compared with an experimental value 0.37 times as large.

The delocalization of the meso-bonding state is also responsible for the hyperfine tensor being virtually isotropic. We find only 2.2% (4.1%) of the deep level charge density is contained in each sulfur p (s) orbital which leads to an anisotropy of the hyperfine tensor too small to have been observed,

$$(A_{//} - A_{\perp})/(A_{//} + 2A_{\perp}) = 0.015.$$

The predicted energy levels and s-orbital wave-function coefficients, γ_s, of all substitutional nearest-neighbor pairs of S, Se, and Te are given in Table I [13,14]. We hope that experimental studies of Si doped with two or more chalcogens will test these predictions.

Recently Ren et al. [15] have shown that the hyperfine tensors of deep levels of definite symmetry associated with impurities on a specific site depend only on the energy of the deep level and vary little for deep levels in the band gap of Si. Schirmer and Scheffler [16] have produced a beautiful phenomenological theory of g-factors for deep levels using similar ideas, and find that the observed g-factors do not depend on the character of the defect states involved. In particular, they find that the g-factor for $(S,S)^+$ is well-described by the same theory as the g-factor for S^+, Se^+, and Te^+. Thus we wondered if the Schirmer-Scheffler hypothesis for g-factors might apply to the hyperfine interactions as well: that the hyperfine interactions for $(S,S)^+$ might be well-described by the theory of hyperfine interactions for the isolated defects S^+, Se^+, and Te^+. It does not (Fig. 3 [17]), presumably because the meso-bonding T_2-derived wave-function of the $(S,S)^+$ pair is entirely different from that of the A_1 isolated-S^+ level. We also predict that the hyperfine tensor of $(Te,Te)^+$ will be quite different from that of Te^+ — a point needing experimental investigation.

TABLE I. Predicted energy levels E (in eV) and wave-function coefficients γ ($\gamma=\alpha n$, see Table I of Ref. [13] or Ref. [14]) for (X,Y) substitutional paired chalcogen defects in Si. E_v and E_c are the valence band and conduction band edges of Si. In the theoretical model the band gap is $E_c-E_v=1.17$ eV.

	Meso-bonding			Totally antibonding		
(X,Y)	E_c-E	$\gamma_s(X)$	$\gamma_s(Y)$	$E-E_v$	$\gamma_s(X)$	$\gamma_s(Y)$
$(S,S)^+$	0.42	0.20	0.20	0.12	0.30	0.30
$(S,Se)^+$	0.41	0.21	0.20	0.16	0.29	0.30
$(S,Te)^+$	0.35	0.17	0.25	0.28	0.32	0.29
$(Se,Se)^+$	0.40	0.20	0.20	0.20	0.30	0.30
$(Se,Te)^+$	0.33	0.17	0.25	0.31	0.33	0.29
$(Te,Te)^+$	0.29	0.21	0.21	0.45	0.32	0.32
$(S,S)^0$	0.28	0.21	0.21	0.44	0.32	0.32
$(S,Se)^0$	0.28	0.23	0.21	0.44	0.31	0.32
$(S,Te)^0$	0.24	0.18	0.25	0.53	0.34	0.30
$(Se,Se)^0$	0.27	0.21	0.21	0.48	0.32	0.32
$(Se,Te)^0$	0.23	0.17	0.26	0.55	0.34	0.30
$(Te,Te)^0$	0.20	0.22	0.22	0.65	0.33	0.33

The calculations presented in this paper employ the Hjalmarson et al. [10] theory of deep levels, the Vogl et al. [18] theory of electronic structure, the Ren at al. [15] theory of hyperfine interactions, and the theory of paired defects [12]. The details of the paired-defect theory can be found elsewhere [12,19]. The present work is different in only one significant way: the choice of the (diagonal) defect potential matrix, in the localized sp^3s^* basis, at the X-site $V\{X\}=(V_s\{X\},V_p\{X\},V_p\{X\},V_p\{X\},0)$ for the (S,X) pairs. The pair has a matrix V{S} on the S-site and V{X} on the neighboring X-site. $V_s\{S\}$ for S in the $(S,S)^+$ complex is chosen to be the average of the values needed to reproduce the observed [7] S^+ and S^0 A_1-symmetric isolated-defect levels — and so corresponds to an isolated-S defect with a charge of +1/2 (so that the paired defect will correspond to a charge of +1). The p potential $V_p\{S\}$ is taken from Refs. [10] and [18] in terms of atomic energy differences, but is reduced by the same empirical factor found here for the s potential, 0.88. For the neutral center $(S,S)^0$, the potential V_s used is that which reproduces the observed A_1-symmetric level of S^0. V{Se} and V{Te} were similarly determined. As a result, the total defect potential of a paired defect such as $(S,S)^+$ represents a pair with a single net positive charge. Other than this determination of the defect potential matrix, the calculation is as described elsewhere [12,19], and contains no free parameters. A complete discussion of the theory of paired substitutional defects will be published shortly.

We hope that this work will stimulate experiments testing the theory of deep levels associated with paired chalcogen impurities in Si.

Acknowledgments — We are grateful to the U. S. Army Research Office for their support of this work (Contract No. DAAG29-81-K-0068). We thank the physics department cf the University of Illinois at Urbana for providing the atmosphere in which the initial ideas leading to this work were developed.

ea.
by

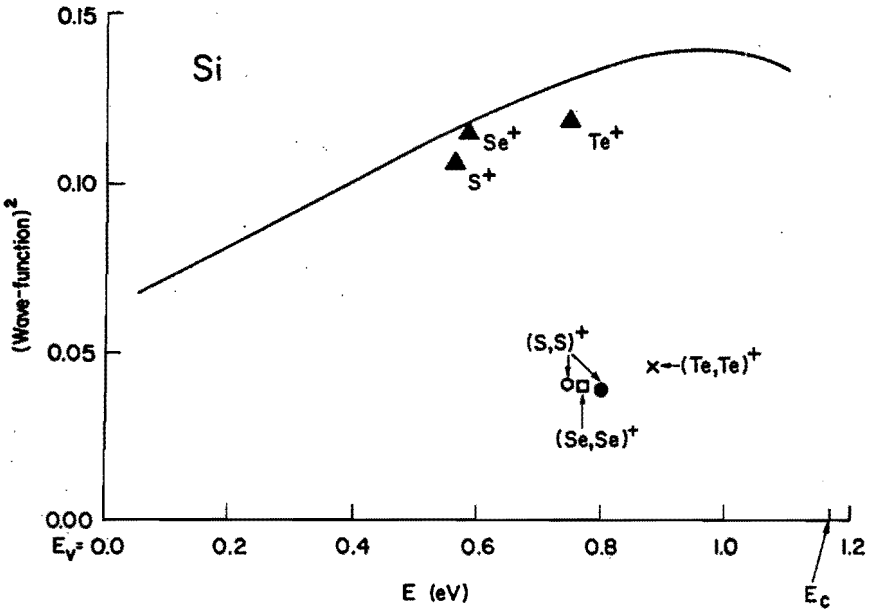


Fig. 3. The square of the isotropic part of the deep level wave-function (solid line), $|\langle A_1,0,1|\psi\rangle|^2$ of Ren et al. [15] versus deep trap energy relative to the valence band maximum for substitutional isolated-chalcogen defects in Si. When multiplied by the atomic charge density $|\psi_{n,s}(0)|^2$, this is proportional to the hyperfine interaction A. The data for S^+, Se^+, and Te^+ [17] obtained using Hartree-Fock charge densities are denoted by closed triangles. The corresponding wave-functions-squared for $(S,S)^+$, $(Se,Se)^+$, and $(Te,Te)^+$ obtained theoretically are denoted by an open circle, an open square, and an ×, respectively. The $(S,S)^+$ datum [1] is denoted by a closed circle.

REFERENCES

[1] G. W. Ludwig, Phys. Rev. 137, A1520 (1964).

[2] W. E. Krag & H. J. Zeiger, Phys. Rev. Lett. 8, 485 (1962); W. E. Krag, W. H. Kleiner, H. J. Zeiger & S. Fischler, J. Phys. Soc. Jpn. Suppl. 21, 230 (1966).

[3] R. O. Carlson, R. N. Hall & E. M. Pell, J. Phys. Chem. Solids 8, 81 (1959).

[4] C. T. Sah, L. L. Rosier & L. Forbes, Appl. Phys. Lett. 15, 316 (1969).

[5] D. L. Camphausen, H. M. James & R. J. Sladek, Phys. Rev. B2, 1899 (1970).

[6] O. Engstrom & H. G. Grimmeiss, J. Appl. Phys. 47, 4090 (1976).

[7] H. G. Grimmeiss, E. Janzén & K. Larsson, Phys. Rev. B25, 2627 (1982); H. G. Grimmeiss, E. Janzén & B. Skarstam, J. Appl. Phys. 51, 4212 (1980).

[8] The assignment of the labels A, B, C, and D to the energy levels below the conduction band edge by 0.61, 0.31, 0.37, and 0.19 eV respectively is not universal in the literature. We use the labelling of the very recent literature of Ref. [7] for the A and B assignments.

[9] S. D. Brotherton, M. J. King & G. J. Parker, J. Appl. Phys. 52, 4649 (1981).

[10] H. P. Hjalmarson, P. Vogl, D. J. Wolford & J. D. Dow, Phys. Rev. Lett. 44, 810 (1980).

[11] Recall that the S deep level is an antibonding state, with much of its charge on adjacent Si sites, not on the S site. Because of this, the linear combination of orbitals that is totally antibonding forms the lower energy state.

[12] O. F. Sankey & J. D. Dow, Phys. Rev B26, 3243 (1982); O. F. Sankey, H. P. Hjalmarson, J. D. Dow, D. J. Wolford & S. G. Streetman, Phys. Rev. Lett. 45, 1656 (1980).

[13] G. D. Watkins & J. W. Corbett, Phys. Rev. 134, A1359 (1964).

[14] J. R. Niklas & J. M. Spaeth, Solid State Commun. 46, 121 (1983).

[15] S. Y. Ren, W. M. Hu, O. F. Sankey & J. D. Dow, Phys. Rev. 26,951 (1982).

[16] O. F. Schirmer & M. Scheffler, J. Phys. C. 15, L645 (1982).

[17] H. G. Grimmeiss, E. Janzén, H. Ennen, O. Schirmer, J. Schneider, R. Wörner, C. Holm, E. Sirtl & P. Wagner, Phys. Rev. B24, 4571 (1982).

[18] P. Vogl, H. P. Hjalmarson & J. D. Dow, J. Phys. Chem. Solids 44, 365 (1983).

[19] O. F. Sankey & J. D. Dow, to be published.

Deep levels associated with impurities at the bond-centered interstitial site in Si

Ward L. Johnson
Department of Physics, University of Illinois at Urbana–Champaign, 1110 West Green Street, Urbana, Illinois 61801

Otto F. Sankey
Department of Physics, University of Illinois at Urbana–Champaign, 1110 West Green Street, Urbana, Illinois 61801 and Department of Physics, Arizona State University, Tempe, Arizona 85287*

John D. Dow
Department of Physics, University of Illinois at Urbana–Champaign, 1110 West Green Street, Urbana, Illinois 61801 and Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556*

(Received 16 September 1983)

The trends in the deep energy levels of impurities occupying the bond-centered interstitial site in Si are predicted. The theory is compared with experiments for boron, and reasonable agreement is found for the energy and the wave-function component on the interstitial site; however, some disagreement with the data for the B deep-trap wave function's amplitude on the adjacent Si sites remains—indicating that the B interstitial may not lie near the bond-centered site as supposed and suggesting the need for further study of B in Si.

In this paper we report calculations of the chemical trends of deep energy levels and wave functions associated with interstitial impurities at the bond-centered site in Si.[1] This work complements earlier studies of tetrahedral-site interstitials by Sankey and Dow[2] and follows the general approach to deep impurity levels established by Hjalmarson *et al.*[3] and Vogl *et al.*[4]

With the notable exception of the work of Weigel,[5] previous studies of interstitial impurities in Si (Refs. 6–8) have considered only single impurities without devoting a great deal of attention to the relationships of the energy levels of one impurity to another. Weigel's systematic studies of Al, Si, P, B, C, and N were performed on a cluster using extended Hückel theory.[9] One disconcerting aspect of his results is a silicon band gap of 7.5 eV; moreover, in the case of substitutional defects, Hückel theory is known to give the T_2 (p-like) energy levels below the A_1 (s-like) levels—contrary to the data.[10] Therefore, we felt that it would be appropriate to perform an independent study of the chemical trends for the bond-centered interstitial impurities, using a Green's-function approach, which yields the correct band gap for Si and the correct A_1-T_2 ordering for substitutional defect levels.[3] Our results confirm the general trends found by Weigel, but produce a factor-of-5 less variation in the interstitial energy levels as one goes from Al to P.

Our calculations follow the general scheme of Sankey and Dow[2] developed for tetrahedral-site interstitials. However, we do not iterate our calculations to self-consistency since we find that most impurities at the bond-centered site (in contrast to the tetrahedral interstitial site) produce neutral deep levels in the gap. We have included only nearest-neighbor interactions between the host atoms and the interstitial atom because the second-nearest neighbors are more than 2.5 times as distant from the interstitial as the nearest-neighbor Si. Lattice relaxation around the defect is neglected, since we are interested primarily in the global chemical trends: the amount and nature of the lattice distortion around the defect should be determined separately for each charge state of each defect. Our deep levels are obtained for neutral defects and the levels of charged defects are obtained by adding (removing) an electron to the lowest (highest) Pauli-available level; thus the Coulombic charge-state splittings are zero in this model. The resulting theoretical uncertainty in the absolute energies of deep levels should be several tenths of an eV, although the chemical trends in the deep levels should be predicted rather well.

For a zinc-blende lattice, the point group of the bond-centered interstitial is C_{3v}. The diamond structure of Si is a degenerate form of zinc blende in which the "anion" and the "cation" are equivalent; the correct point group for a bond-centered interstitial in a diamond lattice is D_{3d} and has an extra parity quantum number: gerade and ungerade. In this paper we use the C_{3v} notation[11] and also indicate the parity by "g" (even) or "u" (odd). The four sp^3 orbitals centered on the interstitial split into an $A_{1,g}$ s-like orbital, an $A_{1,u}$ p_z- or p_σ-like orbital, and a doubly degenerate E_g p_π-like orbital. Our calculation shows that, for all s- and p-bonded interstitial impurities in Si, only the E_g state may lie in the gap—the $A_{1,u}$ state lies down in the valence band and the $A_{1,g}$ state lies well up in the conduction band. One can see that the $A_{1,u}$ and $A_{1,g}$ levels are removed from the gap by considering the Si-J-Si molecule (See Fig. 1), where J is the interstitial. The two neighboring Si atoms (without J) form two $A_{1,g}$ levels—one a bonding combination of s orbitals and the other a bonding combination of p_σ orbitals. Both of these levels lie in the valence band. The interstitial's s orbital (of $A_{1,g}$ symmetry) couples with these two Si $A_{1,g}$ levels and is repelled upward into the conduction band. Similarly, two $A_{1,u}$ conduction-band levels are formed from anti-

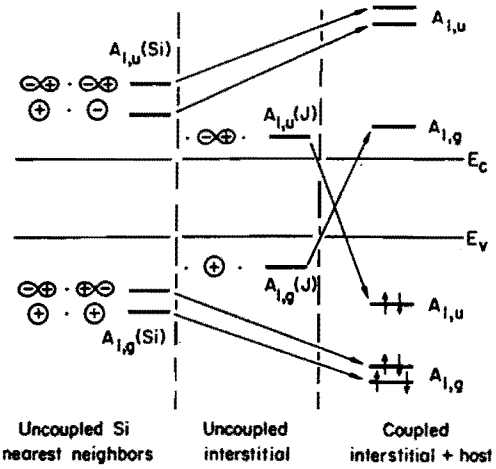


FIG. 1. Schematic illustration showing the crossing of the $A_{1,g}$ and $A_{1,u}$ interstitial levels as a result of the interstitial-host coupling. Only A_1 levels are shown.

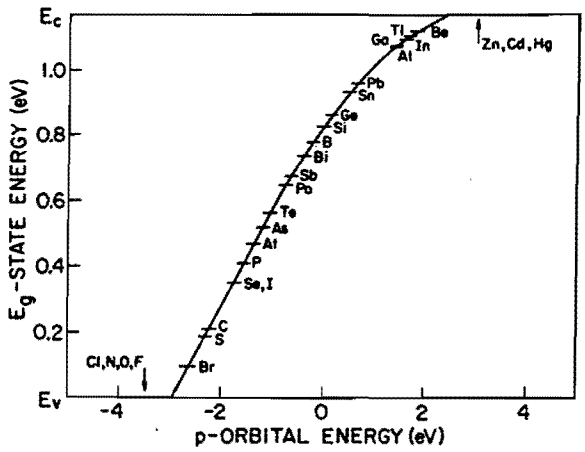


FIG. 2. Bound-state energies of deep levels associated with bond-centered interstitials as a function of the p-orbital energy in the solid of the impurity. The zero of energy for the deep levels is the valence-band maximum; the zero of p-orbital energy is the value for Si. The predicted energy levels for the various s- and p-bonded impurities are indicated on the curve. The differences in p-orbital energies "in the solid" were obtained from the empirical rule of 0.6 times the difference of atomic p-orbital energies (Ref. 4).

bonding combinations of the s and p_σ orbitals of the neighboring two Si atoms. The interstitial's p_σ orbital (of $A_{1,u}$ symmetry) couples with the Si conduction-band $A_{1,u}$ levels and is repelled downward into the valence band. Thus, by being interstitial, the impurity has in effect reversed the ordering of its s and p electrons: the p-like $A_{1,u}$ isolated-interstitial level lies below the s-like $A_{1,g}$ level when the interstitial and the host are coupled.

The p_π orbitals on the neighboring Si are polarized perpendicular to the "molecular" axis and linear combinations of these can be constructed with either E_u or E_g symmetry, with respect to the bond-centered site. The interstitial's p_π orbital couples with the Si p_π-like E_g levels and gives rise to the E_g level which falls in the gap. (E_u levels are unaffected by the interstitial atom and remain in the bands.)

Our predictions for the E_g-symmetric ($p_{\pi,g}$-like) deep levels at the bond-centered interstitial site in Si are given in Fig. 2, and show similar chemical trends to those found for Al, Si, B, P, C, and N by Weigel: (from highest to lowest) Hg, Cd, and Zn (resonances in the conduction band), Be, Tl, In, Ga, Al, Pb, Sn, Ge, Si, B, Bi, Sb, Po, Te, As, At, P, Se, I, C, S, and Br (in the gap), with Cl, N, O, and F giving levels in the valence band.

In our model for interstitials in Si, Al is only 0.86 eV above C, in contrast to Weigel's result which places Al more than 4 eV higher than C. The deep levels of symmetries other than E_g, namely, $A_{1,u}$ and $A_{1,g}$, all lie well outside the gap for all interstitials.

The natural occupation of the E_g deep level is zero electrons for group-II impurities, one electron for group-III, two electrons for group-IV, three electrons for group-V, and four electrons for group-VI impurities. Thus the neutral defects of Pb, Sn, Ge, Si, and C at the bond-centered site produce a one-electron level occupied by two electrons. Bi, Sb, As, and P have three electrons in the E level; nitrogen produces a valence-band resonance occupied initially by three electrons and one hole in this model, and so donates the hole to the top of the valence band, forming a shallow acceptor (provided the extra Coulomb repulsion energy of the fourth electron does not restore the N^- deep level to the gap). Po, Te, Se, S, and O have four electrons in the deep level. At, I, Br, Cl, and F fill the deep level with four electrons and produce a shallow donor electron near the conduction-band edge as well. Beryllium yields an empty electron trap, whereas Hg, Cd, and Zn yield an empty deep resonance above the conduction-band edge. Of course it is not known which, if any, of these impurities are metastable or stable at the bond-centered interstitial site. (Partially occupied E levels are Jahn-Teller unstable.) Moreover, these predictions of absolute level positions may be in error by ~ 0.5 eV, and should be interpreted with this large uncertainty in mind.

Watkins[1] has studied the neutral B^0 interstitial thoroughly and has proposed three different models of the interstitial geometry—one model is the bond-centered interstitial with a significant distortion from the bond-centered site. We predict a level position of $E_c - 0.39$ eV for B^0 undistorted from the bond-centered site in remarkably good agreement with the experimental estimate of $E_c - 0.15$ (Ref. 1) to -0.13 eV (Ref. 12), where E_c is the energy of the conduction-band edge.

We have also calculated the spin-resonance hyperfine parameters a_j and b_j (Ref. 13) and the wave-function coefficients α_j, β_j, and η_j discussed by Watkins.[1] Our results for undistorted B^0 are compared with his experimental results in Table I. Because our theoretical B^0 is not distorted from the bond-centered interstitial site, we have $\alpha_j = 0$ for the theory (the E_g-symmetric state is purely p_π like), and no distinction between what Watkins terms Si1 and Si2. The agreement between theory and data is good, but not excellent: the theory places 25% of the charge on the B atom, versus 33% experimentally, this is reasonable agreement considering the simplicity of the theory and the fact that we have not allowed our B to relax from the bond-centered site. At this point, it would appear that the

TABLE I. Energy (in eV) and amplitude of the E_g state deep-level wave function η_j on the B interstitial and its neighbors Si1 and Si2, and the relative amplitudes of the s and p orbitals, α_j and β_j, respectively. Since the theory considers the undistorted bond-centered site, the s-orbital contribution α_j is zero on all three sites (for E_g symmetry), and the amplitudes on the two silicon atoms are the same.

		Energy		
Theory		$E_c-0.39$		
Experiment		$E_c-0.15$,[a]	$E_c-0.13$[b]	
Amplitudes:		α_j	β_j	η_j
B	Theory	0.0	1.0	0.50
	Experiment[a]	0.17	0.98	0.57
Si1	Theory	0.0	1.0	0.05
	Experiment[a]	0.37	0.93	0.39
Si2	Theory	0.0	1.0	0.05
	Experiment[a]	0.14	0.99	0.44

[a]Reference 1.
[b]Reference 12.

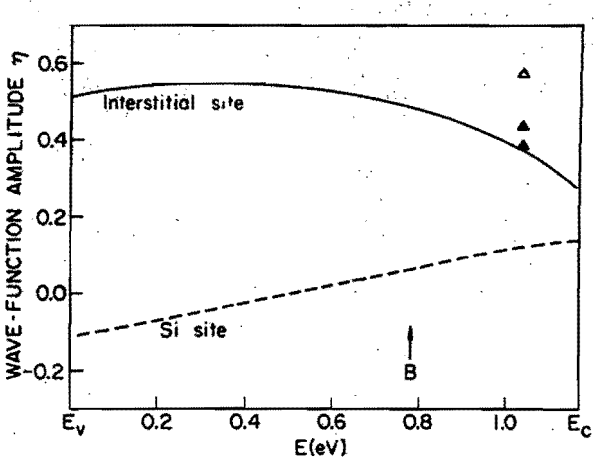


FIG. 3. Amplitudes of the E_g-symmetric wave function on the interstitial site (solid line) and on one of the neighboring silicon atom sites (dashed line) as a function of the deep-level energy E in the Si band gap. The arrow indicates the predicted energy position of the B^0 interstitial. The solid triangles are the B^0-interstitial data (Ref. 1) for the wave-function amplitudes on the two nearest-neighbor Si atoms, and the open triangle represents the data for the B site amplitude. The relative signs of the amplitudes cannot be determined experimentally, but have all been taken to be positive for deep levels near the conduction-band edge.

theory is in satisfactory agreement with the data and that the assignment of B to a bond-centered site, or one slightly distorted from it, is reasonable.

However, the theory predicts that only 0.3% (see Fig. 3) of the B^0 deep level's charge should be found on each of the adjacent Si sites, whereas experiment finds over an order of magnitude more, 15–19% for η^2 (see Table I). Such a discrepancy is alarming, because this basic model has successfully obtained the observed wave-function amplitudes for deep levels associated with substitutional S^+, Se^+, and Te^+ in Si (Ref. 14) and Al^{2+} at the tetrahedral interstitial site in Si.[2] We have examined the theory and tried different tight-binding host Hamiltonians[15] to determine if this result is an artifact of the specific model we have chosen. We do not believe that it is. We conclude, in agreement with earlier work,[16] that the observed large charge density on the adjacent Si atoms is associated with the fact that the neutral B defect is not on the bond-centered interstitial site—either because the lattice has distorted somewhat (as is known to be the case) or because the assignment of the bond-centered site as the parent site is incorrect. However, considerable distortion may be necessary to cause 15% of the deep trap's charge to reside on each of the adjacent Si atoms. Hence, the present calculations indicate a need for a thorough theoretical study of the configuration coordinates of B in Si, and the dependence of the interstitial B deep level and wave function on the position of the B atom in the unit cell.

It is interesting to note that the theoretical B^0 deep level is almost nonbonding with its neighboring Si atoms—an effect noted for interstitials on tetrahedral sites as well.[2] Perhaps this nonbonding character, combined with the reversal of ordering of s and p electrons discussed above, makes B especially vulnerable to lattice distortion and may play a role in its becoming a "negative-U" center.[8,12,17]

We gratefully thank the U.S. Army Research Office (Contract No. DAAG29-81-K-0068) for generous support of this research, and we acknowledge stimulating conversations with M. A. Bowen and A. V. Granato.

APPENDIX

We briefly describe the calculation of bound energy levels of an interstitial at the bond-centered site. This calculation follows the same general procedure described in more detail by Sankey and Dow[2] for the case of tetrehedral-site interstitials.

The Hamiltonian of the host crystal plus the interstitial, in a localized-orbital basis, is given by

$$H=\begin{bmatrix} H_0 & W \\ W^\dagger & H^J \end{bmatrix}, \tag{A1}$$

where H_0 is the Si perfect-crystal Hamiltonian, H^J is the Hamiltonian of the uncoupled interstitial, and W couples the interstitial atom with the host crystal. H_0 is taken to be the empirical $10N\times10N$ sp^3s^* tight-binding matrix of Vogl *et al.*,[4] where N is the number of unit cells. The Hamiltonian of the interstitial atom H^J in an sp^3 basis is a diagonal 4×4 matrix with matrix elements ϵ_s, ϵ_p, ϵ_p, and ϵ_p—the s and p orbital energies. Only nearest-neighbor interactions with the interstitial are included in the calculation, so the only nonzero submatrix involving the interstitial is 12×12 (four orbitals on the interstitial and on each of its two neighbors). The elements of W are taken to be the corresponding elements of the perfect-crystal tight-binding Hamiltonian,[4] scaled with distance according to Harrison's[18] d^{-2} rule, where d is the interstitial-host bondlength.

The interstitial energy levels are given by

$$\det[1-G_{\rm ref}(E)V]=0\ , \tag{A2}$$

where $G_{\rm ref}(E)=(E-H_{\rm ref})^{-1}$ is the Green's function for a (reference) Si interstitial coupled to the host crystal. $H_{\rm ref}$ is the Hamiltonian of Eq. (A1) with a Si atom in the interstitial position and V is the difference between the actual interstitial Hamiltonian and the reference: $H=H_{\rm ref}+V$. The numerical values used for the on-site matrix elements of the reference interstitial are listed in Table II.[19]

In the case of a bond-centered interstitial in Si, the interstitial occupies a site of D_{3d} symmetry. Thus there will be eigenstates of $A_{1,u}$ symmetry, of $A_{1,g}$ symmetry, and of twofold-degenerate E_g symmetry. The symmetrized on-site basis functions of the interstitial are

$$|A_{1,g}\rangle=|s(J)\rangle\ ,$$

$$|A_{1,u}\rangle=|p_z(J)\rangle\ ,$$

$$|E_g,x\rangle=|p_x(J)\rangle\ ,$$

and

$$|E_g,y\rangle=|p_y(J)\rangle\ , \tag{A3}$$

where the z axis is defined to lie along the Si-J-Si axis, and the atom is indicated in the parentheses.

The determinantal equation (A2) reduces to a set of equations of the Koster-Slater type:[20]

$$\langle\lambda|G_{\rm ref}(E)|\lambda\rangle=\langle\lambda|V|\lambda\rangle^{-1}\ , \tag{A4}$$

where $|\lambda\rangle$ is one of the basis states listed above in Eq. (A3). The diagonal elements of the reference Green's function are given by

$$\langle\lambda|G_{\rm ref}(E)|\lambda\rangle=[E-\langle\lambda|H^J_{\rm ref}|\lambda\rangle-\Sigma_\lambda(E)]^{-1}\ , \tag{A5}$$

where we have

TABLE II. Matrix elements (in eV) used in the calculation, using the notation of Ref. 19. These matrix elements were obtained by scaling the bulk Si parameters of Ref. 4 by d^{-2} (Ref. 18).

$W_{sso}=\langle s(\mathrm{Si})\|H\|s(J)\rangle$	$=-8.300$
$W_{spo}=\langle s(\mathrm{Si})\|H\|p_\sigma(J)\rangle$	$=9.924$
$W_{pp\pi}=\langle p_\pi(\mathrm{Si})\|H\|p_\pi(J)\rangle$	$=-2.860$
$W_{pp\sigma}=\langle p_\sigma(\mathrm{Si})\|H\|p_\sigma(J)\rangle$	$=10.864$
$\epsilon_s(\mathrm{Si})=\langle s(\mathrm{Si})\|H\|s(\mathrm{Si})\rangle$	$=-4.200$
$\epsilon_p(\mathrm{Si})=\langle p(\mathrm{Si})\|H\|p(\mathrm{Si})\rangle$	$=1.715$

$$\Sigma_\lambda(E)=\langle\lambda|W^\dagger G_0(E)W|\lambda\rangle\ . \tag{A6}$$

The only matrix elements which we have not yet specified are those of the perfect-crystal Green's function $G_0(E)=(E-H_0)^{-1}$. These are computed from the perfect-crystal tight-binding Hamiltonian H_0 using the special-points method.[21]

For the twofold-degenerate E_g level, Eqs. (A4)–(A6) reduce to

$$E-\epsilon_p(\mathrm{Si})-2W^2_{pp\pi}[\langle p_x(\mathrm{Si1})|G_0|p_x(\mathrm{Si1})\rangle+\langle p_x(\mathrm{Si1})|G_0|p_x(\mathrm{Si2})\rangle]=V_p\ . \tag{A7}$$

Here the two Si atoms are denoted Si1 and Si2.

*Permanent address.

[1]G. D. Watkins, Phys. Rev. B 12, 5824 (1975).

[2]O. F. Sankey and J. D. Dow, Phys. Rev. B 27, 7641 (1983).

[3]H. P. Hjalmarson, P. Vogl, D. J. Wolford, and J. D. Dow, Phys. Rev. Lett. 44, 810 (1980).

[4]P. Vogl, H. P. Hjalmarson, and J. D. Dow, J. Phys. Chem. Solids 44, 365 (1983).

[5]C. Weigel, in *International Conference on Radiation Effects in Semiconductors, Nice, 1978*, edited by J. H. Albany (IOP, London, 1979), p. 186.

[6]V. A. Singh, C. Weigel, J. W. Corbett, and L. M. Roth, Phys. Status Solidi B 81, 637 (1977).

[7]S. P. Singhal, Phys. Rev. B 5, 4203 (1972).

[8]G. A. Baraff, M. Schlüter, and G. Allan, Phys. Rev. Lett. 50, 739 (1983).

[9]F. P. Larkins, J. Phys. C 4, 3065 (1971); R. P. Messmer and G. D. Watkins, Phys. Rev. B 7, 2568 (1973); M. Astier, N. Pottier, and J. C. Bourgoin, *ibid.* 19, 5265 (1979).

[10]M. Lannoo, Phys. Rev. B 25, 2987 (1982).

[11]M. Tinkham, *Group Theory and Quantum Mechanics* (McGraw-Hill, New York, 1964).

[12]R. D. Harris, J. L. Newton, and G. D. Watkins, Phys. Rev. Lett. 48, 1271 (1982).

[13]The hyperfine tensor has parallel, $A^j_{||}=a_j+2b_j$, and perpendicular, $A^j_\perp=a_j-b_j$, components. The quantities a_j and b_j are proportional to the s- and p-orbital spin densities, respectively, on atom j.

[14]S. Y. Ren, W. M. Hu, O. F. Sankey, and J. D. Dow, Phys. Rev. B 26, 951 (1982).

[15]D. A. Papaconstantopoulos and E. N. Economou, Phys. Rev. B 22, 2903 (1980).

[16]A. Mainwood, Solid-State Electron. 21, 1431 (1978), and P. Masri, A. H. Harker, and A. M. Stoneham, J. Phys. C 16, L613 (1983), have performed total-energy calculations and have reached similar conclusions.

[17]J. R. Troxell and G. D. Watkins, Phys. Rev. B 22, 921 (1980).

[18]W. A. Harrison, *Electronic Structure and the Properties of Solids* (Freeman, San Francisco, 1980).

[19]J. C. Slater and G. F. Koster, Phys. Rev. 94, 1498 (1954).

[20]G. F. Koster and J. C. Slater, Phys. Rev. 95, 1167 (1954).

[21]D. J. Chadi and M. L. Cohen, Phys. Rev. B 8, 5747 (1973).

FILE: [dow.manus]milwau.rno 18 November 1983 15:45:33

Intrinsic and extrinsic surface electronic states of semiconductors

John D. Dow

Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556

and

Roland E. Allen

Department of Physics, Texas A&M University, College Station, Texas 77843

and

Otto F. Sankey

Department of Physics, Arizona State University, Tempe, Arizona 85281

Abstract

Recent theoretical work on the physics of surface deep impurity states, intrinsic surface states, and surface core exciton states in semiconductors is reviewed. The Schottky barrier ehights of the common semiconductors can be understood in terms of Fermi-level pinning by various surface deep levels associated with native defects or defects produced by surface treatments. The same theoretical framework, which has been successfully applied to bulk deep-level problems, also provides a good account of the physics of intrinsic surface state dispersion relations and surface core exciton states.

make
may

I. Introduction

In the bulk of a tetrahedral semiconductor, a single substitutional s-p bonded impurity or vacancy will ordinarily produce four "deep" levels with energies near the fundamental band gap: one s-like (A_1) and three p-like (T_2) [1]. These deep levels may lie within the fundamental band gap, in which case they are conventional deep levels, or they may lie within either the conduction or the valence band as "deep resonances." A sheet of N vacancies will produce 4N such deep levels — namely, the intrinsic surface state energy bands, which may or may not overlap the fundamental gap (to a good approximation, insertion of a sheet of vacancies is equivalent to creating a surface).

Intrinsic surface states have common underlying physics with deep impurities because they too result from localized perturbations of a semiconductor [2], and so their energies can be relatively easily predicted by extending to surfaces ideas developed by Hjalmarson, Vogl, Wolford, et al. [1] for the deep impurity problem. This has been done by several authors [3][4][5][6][7][8][9][10][11][12], most notably by Allen and co-workers [13][14][15][16][17].

Extrinsic and native-defect surface states also are governed by similar physics, and are especially interesting in the light of the Schottky barrier problem: Bardeen showed that modest densities of surface states on a semiconductor can "pin" the Fermi level [18], forming a Schottky barrier. The bulk Fermi energies of the semiconductor, the metal, and the semiconductor surface must align (Fig. 1). If the semiconductor is heavily doped n-type, the surface Fermi energy is the lowest empty surface state. The bands bend to

accomodate this alignment of Fermi levels, forming the Schottky barrier. Thus the Schottky barrier height is the binding energy of the lowest naturally empty surface state, relative to the conduction band edge. In 1976 Spicer et al. proposed that the Bardeen surface states responsible for pinning the Fermi energy are due to native defects [18][19][20][21][22].

Surface core excitons are similar to surface defect states, as can be seen by using the optical alchemy approximation [23] or the Z+1 rule [24]. Consider core excitation of a Ga atom at the surface of GaAs; the radius of the core hole is sufficiently small that the hole can be assumed to have zero radius (i.e., the hole is equivalent to an extra proton in the nucleus). Thus the core-excited electron feels the potential of an atom whose atomic charge Z is greater than that of Ga by unity, namely Ge. Thus the Ga core exciton spectrum is approximately the same as the spectrum of a Ge impurity on a Ga site. Hence the core exciton states in semicondutors can be either "shallow" (Wannier-Mott excitons) or "deep" (Hjalmarson-Frenkel excitons), as is the case for impurity states. The deep Hjalmarson-Frenkel excitons are similar to the surface deep levels associated with impurities.

In this paper. we show that the physics of deep impurity levels, intrinsic surface states, surface impurity states, Schottky barriers, and Hjalmarson-Frenkel core excitons are all similar.

II. Deep impurity levels at the surface: Schottky barriers and Fermi-level pinning

The basic physics of most Schottky barriers can be explained in terms of the Fermi-level pinning idea of Bardeen [18]. Stated in a slightly oversimplified form for a degenerately doped semiconductor at zero temperature, the Fermi energies of the metal, the bulk semiconductor, and the semiconductor surface all align in electronic equilibrium. For an n-type semiconductor with a distribution of electronic states at the surface, the Fermi level of the neutral surface is the energy of the lowest states that is not fully occupied by electrons. Electrons diffuse, causing band-bending near the semiconductor surface, until the surface Fermi energy aligns with the Fermi levels of the bulk semiconductor and the metal. This results in the formation of a potential barrier betweeen the semiconductor and the metal, the Schottky barrier (Fig. 1). For an n-type semiconductor, the Schottky barrier height is essentially the energy separation between the surface state that is the Fermi level and the conduction band edge. For a p-type semiconductor, the barrier height is the energy of the highest occupied electronic state of the neutral surface, relative to the valence band maximum. Thus the problem of determining Schottky barrier heights is reduced to obtaining the energy levels of the surface states responsible for the Fermi-level pinning.

In his original article, Bardeen focussed his attention on intrinsic semiconductor surface states as the most likely candidates for Fermi-level pinning. But he also pointed out that deep levels in the gap associated with impurities or native surface defects could also be responsible for the phenomenon.

Following Bardeen's work, a major advance occurred as a result of the experiments of Mead and Spitzer [25] who determined the Schottky barrier heights of many semiconductors, both n-type and p-type. Most of those old data have been confirmed by modern measurements taken under much more favorable experimental conditions.

However, after this work, the Schottky barrier problem was widely regarded as understood [26] in terms of concepts quite different from Fermi-level pinning.

In recent years Spicer and co-workers have revived the Fermi-level pinning model and have argued that the pinning is accomplished by native defects at or near the surface. Their picture is that during the deposition of the metal native defects are created at or near the semiconductor/metal interface, and that these semiconductor surface defects produce deep levels in the band gap that are responsible for Fermi-level pinning.

Spicer's viewpoint has been contested by Brillson and co-workers [27], who have emphasized the importance of chemical reactivity on barrier height. The Brillson viewpoint gains support from the observation of well-defined chemical trends in the variation of barrier height with the heat of reaction of the metal/semiconductor interface, as shown for n-InP by Williams et al. [28][29][30] (Fig. 2). (We believe that the Spicer and Brillson viewpoints can be reconciled.)

Daw, Smith, Swarts, and McGill [31] have proposed that free surface vacancies account for some of the observed Schottky barrier heights in III-V semiconductors. Allen and co-workers have argued that antisite defects

[32][33][34][35][36] "sheltered" [37] at the surface pin the Fermi energy for most Schottky barriers between III-V semiconductors and non-reactive metals, but that vacancies become the dominant pinning defect when the metal is reactive [36]. Thus the Brillson reactivity picture can be unified with the Spicer Fermi-level pinning picture: the chemical reaction merely changes the dominant pinning defect. The experimental results of Mead and Spitzer [25], Wieder [38][39][40], Williams [28][29][30], Mönch [41][42][43][44][45], their co-workers, and many others support this general viewpoint.

Moreover, the connection between the Schottky barriers formed at Si interfaces with transition metal silicides and the barriers between III-V semiconductors and metals appears to be provided by the recent work of Sankey et al. [46]: Fermi-level pinning can account for the silicide data as well. Thus a single unifying picture of Schottky barrier heights in III-V and homopolar semiconductors appears to be emerging. And although this Fermi-level pinning picture is no doubt oversimplified, it does provide a simple explanation of the first-order physics determining Schottky barrier heights, and how the physics changes when the dominant defect switches as a result of chemical reactivity.

It appears unlikely, however, that the Fermi-level pinning mechanism of Schottky barrier formation is universal. Layered semiconductors appear <u>not</u> to exhibit Fermi-level pinning, but rather seem to obey the original Schottky model [30]. This is probably because the layered semiconductors' surfaces are relatively impervious to defects and do not have defect levels in the band gap.

The Fermi-level pinning mechanism of Schottky barrier formation has the most advocates for III-V semiconductors such as GaAs and InP. However, even for these materials there are other proposed mechanisms for Schottky barrier formation, most notably those of Freeouf [47] and Ludecke [48].

Studies of Si, especially Si/transition-metal silicide interfaces, have focussed on the role of the silicide in Schottky barrier formation [49], in contrast to the studies of III-V's. Thus, prior to the recent work of Sankey et al. [46], it was widely believed that Fermi-level pinning was not responsible for the Schottky barrier at these silicide interfaces.

Thus the present state of the field is that Fermi-level pinning has its advocates for some semiconductors, but is not generally accepted as a universal mechanism of Schottky barrier formation, especially at Si/transition-metal silicide interfaces.

A central point of this paper is the Fermi-level pinning can explain an enormously wide range of phenomena relevant to Schottky barrier formation in III-V semiconductors and in Si — which no other existing model can do. In fact, the authors believe that Fermi-level pinning by native defects is responsible for the Schottky barrier formation in III-V semiconductors and in Si.

Our approach to the problem is simple: we calculate deep levels of defects at surfaces and interfaces, and we use these calculations to interpret existing data in terms of the Fermi-level pinning model. To illustrate our approach, we first consider the Si/transition-metal silicide interface and Fermi-level pinning by dangling bonds, as suggested by Sankey et al. [46].

a) Si/transition-metal silicide Schottky barriers

A successful theory of Si/transition-metal silicide Schottky barrier heights must answer the following questions: (1) How are the Schottky barrier heights at Si/transition-metal silicide interfaces related to those at interfaces of III-V semiconductors with metals and oxides? (2) Why is it that Schottky barrier heights of Si with different transition metals do not differ by ~1 eV, since changes of silicide electronic structure on this scale are known to occur [50]? (3) What is the explanation of the weak chemical trends that occur on a ~0.1 eV scale [50]? (4) Why are the Schottky barrier heights of silicides with completely different stoichiometries, such as Ni_2Si, NiSi, and $NiSi_2$ all equal to within ~0.03 eV? (5) Why are the Schottky barrier heights virtually independent of the silicide crystal structure? (6) Why is it that barriers form with less than a monolayer of silicide coverage? (7) Why do the Schottky barrier heights for n- and p-Si very nearly add up to the band gap of Si? (8) What role do the d-electrons of the transition metal play in Schottky barrier formation?

The answers to all of these questions are simple and straightforward, if one proposes (as Sankey et al. [46] have done) that the Si/transition-metal silicide Schottky barriers are a result of Fermi-level pinning by Si dangling bonds at the Si/transition-metal silicide interface. (1) The Fermi-level pinning idea unifies the Si/transition-metal silicide Schottky barriers with those found for the III-V's. (2) The Schottky barrier heights' independence of the transition-metal silicide comes from the fact that the causative agent, the Si dangling bond, is associated with the Si, and not with the silicide of transition metal. (3) The weak chemical trends in barrier heights occur

because the different transition-metal silicides repel the Si dangling bond wavefunction somewhat differently, causing it to lie slightly more or less in the Si. (4,5) The Schottky barrier heights vary very little with silicide stoichiometry and silicide crystal structure because the Si dangling-bond level is "deep-level pinned" in the sense of Hjalmarson et al. [1]: a large change in defect potential produces only a small change in the deep level responsible for Fermi-level pinning. The transition metal atoms act as inert encapsulants with the electronic properties of vacancies, because their energy levels are out of resonance with the Si. (6) Sub-monolayer barrier formation occurs because the Si dangling-bond defect responsible for the Fermi-level pinning is a localized defect that forms before a full interface is formed. (7) The Schottky barrier heights for n-Si and p-Si add up to the band gap because (in a one-electron approximation) the pinning level associated with the neutral Si dangling-bond at the interface is occupied by one electron, and so can accept either an electron or a hole: it is the surface Fermi level for both electrons and holes — both the lowest partially empty state and the highest partially filled state. (8) The d-electrons of the transition metal atoms play no essential role in the transition-metal silicide Schottky barrier formation, except to determine the occupancy of the Si dangling bond deep level; they are out of resonance with the Si at the interface.

The physics of the Si dangling-bond, Fermi-level pinning mechanism is contained in the very simple model presented by Sankey et al. [46]: to a good approximation, a Si dangling-bond at a Si/transition-metal silicide interface is the same as a vacancy in bulk Si with three of its four neighbors replaced by transition-metal atoms. To illustrate this physics, consider first a vacancy in bulk Si. This defect produces four deep levels near the band gap: a

non-degenerate A_1 or s-like level deep in the valence band (a "deep resonance") and a three-fold degenerate T_2 level in the band gap. The Si dangling bond defect at a Si/transition-metal silicide interface differs for the bulk Si vacancy in two ways: (1) some of the nearest-neighbors of the interfacial vacancy are transition-metal atoms rather than Si atoms; and (2) more distant neighbors are also different atoms at different positions -- but the experimental fact that Schottky barriers form at submonolayer coverages suggests that these differences in remote atoms are unimportant. Thus we can imagine constructing the Fermi-level pinning defect by slowly changing some of the Si atoms adjacent to a bulk Si vacancy into transition-metal atoms (Fig. 3).

To be specific we consider a $Si/NiSi_2$ interface, with a missing Si-bridge atom. Thus (Fig. 4) the Si bond dangles into the vacancy left by the removal of the Si bridge atom; this vacancy is surrounded by one Si atom and three Ni atoms.

How are the Ni atoms different from Si? First, their s and p orbital energies lie well above those of Si. Second, they each have an additional d orbital, with an energy that lies well below the Si s and p orbital energies (and is not terribly relevant here). The very positive Ni s and p energies act as a repulsive potential barrier to electrons, repelling the Si dangling bond electron from their vicinity in the silicide and forcing it to reside almost exclusively in the Si.

The effect of this positive potential barrier due to the Ni-Si difference, as it is turned on slowly in our imagination, is to drive the levels of the bulk vacancy upward in energy. In fact, for Ni, the potential is sufficiently positive to drive the T_2 bulk-Si vacancy level out of the gap into the conduction band. At the same time, the A_1 deep resonance of the Si bulk vacancy is also driven upward. For sufficiently large and positive potential, it pops into the fundamental band gap.

The A_1-derived level cannot be driven all of the way through the gap by the potential though, because an (approximate) level-crossing theorem prevents this. A simple way to see that there is an upper bound within the gap for the perturbed A_1 level is to consider a paired-defect of a vacancy V_{Si} with a neighboring atom X. If the atom X is Si, then the defect levels are the A_1 (s-like) valence band resonance and T_2 (p-like) band gap deep level of the bulk Si vacancy. A_1 and T_2 are not good irreducible representation labels of the (V_{Si},X) pair however; the A_1 level becomes σ-bonded and the T_2 level produces one σ-bonded and two π-bonded orbital, with the σ-bond oriented along the V_{Si},X axis and with the π bonds perpendicular to it. Thus the unperturbed (X=Si) σ levels of the (V_{Si},X) pair are the A_1 and T_2 bulk Si vacancy levels. The interlacing or no-crossing theorem [51] states that a perturbation cannot move a level further than the distance to the nearest unperturbed level. (It applies only approximately here.) Hence no matter how electropositive X is, the (V_{Si},X) level derived from the Si vacancy A_1 level cannot lie above the Si vacancy T_2 level. These considerations for general (V_{Si},X) pairs hold for the specific case of (V_{Si},Ni) pairs, and carry over to the dangling bond defect at the Si/transition-metal silicide interface, which is a vacancy surrounded by three Ni atoms and one Si. Thus the dangling-bond A_1 deep level is "deep-level

pinned" (as distinct from Fermi-level pinned) in the sense of Hjalmarson et al. [1], and is insensitive to even major changes in the nearby transition-metal atoms. To a good approximation, the nearby transition-metal atoms have the same effect as vacancies (which can be simulated [52] by letting the orbital energies of the transition-metal atoms approach $+\infty$, thereby decoupling the atoms from the semiconductor).

Thus the work of Sankey et al. [46] not only provides an explanation of the Si/transition-metal silicide Schottky barriers, it explains why calculations for defects at a free surface often can provide a very good description of the physics of Schottky barriers: the defects at interfaces are "sheltered" [37] or encapsulated by vacancies or by metal atoms that have orbital energies out of resonance with the semiconductor atoms; because of the deep-level pinning, the free-surface defects (which can be thought of as encapsulated by vacancies) have almost the same energies as the actual interfacial defects.

b) III-V Schottky barriers

The Fermi-level pinning story for Si/transition-metal silicides holds for Schottky barriers formed on III-V semiconductors as well. Here we summarize the main predictions of the theory.

The basic approach of the theory was to calculate the energy levels in the band gap of thirty s- and p-bonded substitutional point defects at the relaxed [53] (110) surfaces of III-V semiconductors. With these results in hand, Allen et al. examined Schottky barrier data in the context of Fermi-level pinning and eliminated from consideration all defects that

produced levels considerably farther than ~0.5 eV (the theoretical uncertainty) from the observed pinning levels. Interstitial defects were not considered; they have less of a tendency [54] to exhibit the deep-level pinning that is responsible for the experimental fact that different metals produce similar Schottky barrier heights. Moreover, extended defects were not considered initially, because it is known that paired-defect spectra are intimately related to and similar to isolated isolated-defect spectra [55]. (A more complete theory of Fermi-level pinning by paired defects, especially in GaSb where vacancy-antisite pairs are important, is in preparation.)

For clean semiconductors, the native substitutional defects potentially responsible for the commonly observed Fermi-level pinning are vacancies and antisite defects (anions on cation sites or cations on anion sites).

In GaAs, the defects proposed by Allen et al. [32] as responsible for Fermi-level pinning and Schottky barrier formation are the antisite defects. The cation-on-the-As-site defect accounts for trends with alloy composition of the Schottky barrier heights of n-type $In_{1-x}Ga_xAs$ and $Ga_{1-y}Al_yAs$ alloys (Fig. 5). The Fermi-level pinning of p-InAs [56], which shows quite different alloy dependences [57], is also explained.

This picture of Fermi-level pinning has been confirmed recently by Mönch and associates, who annealed Schottky barriers and showed that the Fermi-level pinning disappeared at the same temperature that the bulk (and presumably also the surface) antisite defect is known to anneal [58].

InP is an even more interesting material, because its Schottky barrier appears to depend on the heat of reaction of the interface [28][29][30]. This can be readily explained [36] however in terms of switching of the dominant Fermi-level pinning defect from an antisite defect for non-reactive metals to a vacancy for reactive metals (Fig. 2).

Moreover, surface treatments are known to alter the Schottky barrier height of n-InP, in a manner that can be easily understood in terms of the theory [36]: Surface treatments with Sn or S produce shallow donor levels associated with Sn_{In} or S_P at the surface, and these levels pin the surface Fermi energy for contacts between n-InP and the non-reactive noble metals. Likewise O and Cl treatments lead to reactions with P that leave P-vacancies, so that the surface Fermi-level of treated n-InP interfaced with non-reactive metals lies near the conduction band edge -- as though the metals were reactive.

Thus the Fermi-level pinning idea appears to provide a simple and unifying understanding of a wide variety of Schottky barrier data in the common semiconductors.

III. Intrinsic surface states

The calculations of surface defect levels for the Schottky barrier problem can be checked by simultaneously evaluating surface state energies and comparing them with the considerable body of available data. The theory underlying surface state calculations is basically the same as that for bulk point defects or surface defects. It is quite simple, and requires only (1) the well-established empirical tight-binding Hamiltonian of the semiconductor

[59] (the matrix elements of the Hamiltonian exhibit manifest chemical trends from one semiconductor to another), and (2) knowledge of the positions of the atoms at the surface. Thus a reliable treatment of the surface states of a semiconductor requires an adequate model of the geometrical structure of the surface. At present, no semiconductor surface structures are beyond controversy [60], but two seem to be rather well accepted; the (110) surface structure of III-V and II-VI semiconductors with the zincblende [53][61][62][63], and the $(10\bar{1}0)$ surface structure of II-VI semiconductors with the wurtzite structure [62]. In particular, (110) zincblende surfaces are characterized by an outward, almost-rigid-rotation relaxation of the anion (e.g., As in GaAs), with the bond between surface anion and surface cation rotating through about 27° (III-V's) or 33° (II-VI's), and with small bond length changes and subsurface relaxations.

a) (110) surfaces of III-V and II-VI zincblende semiconductors

During the past five years, a number of groups have reported experimental and theoretical studies of intrinsic surface states at (110) zincblende surfaces [3][4][5][6][7][8][9][10] [11][12][13][14][15][16][17][18] [64][65][66][67][68][69][70][71][72][73]. In Fig. 6, we show the most recent calculation for the dispersion curves $E(\bar{k})$ at the GaAs (110) surface [14], together with the measured surface state energies fo Williams, Smith, and Lapeyre [65] and of Huijser, van Laar, and van Rooy [66]. The calculation employs the ten-band sp^3s^* empirical tight-binding model of Vogl et al. [59]. The agreement between theory and experiment is excellent. For example, along the symmetry lines $\bar{X}'\bar{\Gamma}$ and $\bar{\Gamma}\bar{X}$ (i.e., the boundary of the surface Brillouin zone), the uppermost branch of observed states appears to be explained by A_5,

the next branch by the overlapping resonances A_4 and A_2', and the three lower branches by A_1', A_3, and C_2. Here "A" and "C" refer to states localized primarily on anion and cation sites, respectively. A detailed comparison with previous theoretical studies of the GaAs (110) surface is given in Ref. [14]. The primary additional features are (i) the states A_1 through A_5 and C_1 through C_4 (in the notation of Ref. [7]) were located as bound states or resonances at all planar wavevector $\bar{k}$ along the symmetry lines of the surface Brillouin zone, and (ii) two "new" resonances, A_1' and A_2' were found. (The branch A_1' was reported in Refs. [5] and [74], but not in the other theoretical studies. The branch A_2' had not been previously reported.) The discovery of this additional resonant structure is apparently due to an improved technique for calculating bound states and resonances -- the "effective Hamiltonian" technique [14].

In Fig. 7, the theoretical dispersion curves of Beres et al. [14] are shown for the (110) surface of ZnSe, together with the measured surface state energies reported by Ebina et al. [11]. Again, the agreement between theory and experiment is quite satisfactory, being a few tenths of an eV near the band gap, and larger for more distant states. Some apparent discrepancies [11] between experiment and previous theory were found to be resolved by a more complete treatment of the resonances, using the approach described above.

Surface state dispersion relations have also been calculated for GaP, GaSb, InP, InAs, InSb, AlP, AlAs, AlSb, and ZnTe [14][15][16][17]. In none of the direct-gap materials were intrinsic surface states found within the band gap. GaP, however, was found to have a band of unoccupied surface states that overlaps the fundamental band gap and extends below the bulk conduction band

edge. This is in accord with the experimental facts: of these semiconductors only GaP has surface states in the gap [18][69][70][71][72]. Of the remaining indirect-gap materials, the theory indicates that intrinsic surface states may be observable near the top of the band gap in the indirect-gap Al-V compounds [16], although the theory is not sufficiently accurate to predict unequivocally that the states will lie within the gap.

b) Si (100) (2×1) intrinsic surface states

After many years of intensive study by numerous groups, there is still controversy over the geometrical structures of the most thoroughly studied semiconductor surfaces: Si (100) (2×1) and Si (111) (2×1). For example, four groups have recently given arguments for antiferromagnetic ordering of Si (111) surfaces [75], whereas Pandey has proposed replacing the conventional buckling model [76][77][78] of Si (111) (2×1) by a (110)-like chain model [79].

In the case of Si (100), arguments have recently been presented [80][81][82] against the (2×1) asymmetric dimer model of Chadi [83]. (In the asymmetric dimer model, adjacent rows of surface atoms dimerize, forming a pattern of paired atomic rows on the surface.) The most telling of these arguments involves the apparent disagreement between angle-resolved photoemission measurements of the surface-state dispersion curves [64][65] and theoretical calculations of these dispersion curves with conventional models of the electronic structure as applied to the asymmetric dimer geometry [83][84].

Very recently, two new calculations have been performed independently with improved models of the electronic structure [85][86]. The same conclusion was reached in both of these studies: the electronic structure calculated for the asymmetric dimer model is in agreement with the measurements. This is illustrated in Fig. 8 (taken from Ref. [48]), where both the theoretical band width of 0.65 eV and the detailed variation with the planar wavevector $\bar{k}$ are seen to be in excellent agreement with the experimental dispersion curves. In addition, there is quite satisfactory agreement between the theoretical surface band gaps and the 0.6 eV gap measured by Mönch et al. [87].

IV. Surface core exciton states

The same calculations that predict native-defect surface deep levels for the Schottky barrier problem also yield surface core exciton energies, because the optical alchemy or Z+1 rule states that the Hjalmarson-Frenkel core exciton energies are the energies of "impurities" that are immediately to the right in the Periodic Table of the core-excited atom [23][24]. Thus core-excited Ga produces a "Ge defect" and core-excited In yields "Sn."

In Figs. 9 and 10, the theoretical exciton energies for the (110) surfaces of the Ga-V and In-V compounds are compared with experiment [88]. Notice that the experimental and theoretical exciton levels for InAs and InSb lie above the conduction band edge, as resonances rather than as bound states. In the present theory this result has a simple physical interpretation: Like a deep impurity state, the Hjalmarson-Frenkel exciton energy is determined primarily by the high-density-of-states regions of the bulk band structure. There is only a small density of states near the low-lying direct conduction

band minimum (corresponding to the Γ-point of the Brillouin zone), but a large density of states near the higher, indirect X minima. Thus the conduction band minimum near Γ has relatively little influence on the position of the exciton.

The surface Hjalmarson-Frenkel core excitons have also been calculated for the (110) surface of ZnSe and ZnTe [89] and are in good agreement with the measurements [90]. We conclude that the present theoretical framework does a good job of explaining the basic physics of the "deep" Hjalmarson-Frenkel core excitons, whether bound states or resonances.

V. Unified picture

Thus one interlocking theoretical framework successfully predicts the correct physics of (1) surface deep impurity levels and Schottky barrier heights, (2) intrinsic surface states, and (3) Hjalmarson-Frenkel core exciton states.

REFERENCES

[1] H. P. Hjalmarson, P. Vogl, D. J. Wolford, and J. D. Dow, Phys. Rev. Letters 44, 810 (1980).

[2] H. P. Hjalmarson, H. Büttner, and J. D. Dow, J. Vac. Sci. Technol. 17, 993 (1980); R. E. Allen and J. D. Dow, J. Vac. Sci. Technol. 19, 383 (1981).

[3] C. Calandra, F. Mannghi, and C. M. Bertoni, J. Phys. C 10, 1911 (1977).

[4] J. D. Joannopoulos and M. L. Cohen, Phys. Rev. B10, 5075 (1974); E. J. Mele and J. D. Joannopoulos, Phys. Rev. B17, 1816 (1978).

[5] D. J. Chadi, J. Vac. Sci. Technol. 15, 631, 1244 (1978); Phys. Rev. B18, 1800 (1978).

[6] J. A. Knapp, D. E. Eastman, K. C. Pandey, and F. Patella, J. Vac. Sci. Technol. 15, 1252 (1978).

[7] J. R. Chelikowsky and M. L. Cohen, Phys. Rev. B20, 4150 (1979).

[8] A. Mazur, J. Pollmann, and M. Schmeits, Solid State Commun. 42, 37 (1982).

[9] A. McKinley, G. P. Srivastava, and R. H. Williams, J. Phys. C 13, 1581 (1980); R. H. Williams et al., to be published.

[10] M. Schmeits, A. Mazur, and J. Pollmann, Solid State Commun. 40, 1081 (1981).

[11] A. Ebina, T. Unno, Y. Suda, H. Koinuma, and T. Takahashi, J. Vac. Sci. Technol. 19, 301 (1981).

[12] F. Manghi, E. Molinari, C. M. Bertoni, and C. Calandra, J. Phys. C 15, 1099 (1982).

[13] R. E. Allen, H. P. Hjalmarson, and J. D. Dow, Surf. Sci. 110, 2625 (1981).

[14] R. P. Beres, R. E. Allen, and J. D. Dow, Solid State Commun. 45, 13 (1983).

[15] R. P. Beres, R. E. Allen and J. D. Dow, Phys. Rev. B26, 769 (1982).

[16] R. P. Beres, R. E. Allen, J. P. Buisson, M. A. Bowen, G. F. Blackwell, H. P. Hjalmarson, and J. D. Dow, J. Vac. Sci. Technol. 21, 548 (1982).

[17] R. P. Beres, R. E. Allen, and J. D. Dow, Phys. Rev. B26, 5702 (1982).

[18] J. Bardeen, Phys. Rev. 71, 717 (1947).

[19] W. E. Spicer, P. W. Chye, P. R. Skeath, C. Y. Su, and I. Lindau, J. Vac. Sci. Technol. 16, 1422 (1979), and references therein.

[20] W. E. Spicer, I. Lindau, P. R. Skeath, and C. Y. Su, J. Vac. Sci. Technol. 17, 1019 (1980), and references therein.

[21] W. E. Spicer, I. Lindau, P. R. Skeath, C. Y. Su, and P. W. Chye, Phys. Rev. Letters 44, 520 (1980).

[22] P. Skeath, C. Y. Su, I. Hino, I. Lindau, and W. E. Spicer, Appl. Phys. Lett. 39, 349 (1981).

[23] J. D. Dow, D. R. Franceschetti, P. C. Gibbons, and S. E. Schnatterly, J. Phys. F 5, L211 (1975) and references therein.

[24] H. P. Hjalmarson, H. Büttner, and J. D. Dow, Phys. Rev. B24, 6010 (1981).

[25] C. A. Mead and W. G. Spitzer, Phys. Rev. A134, 713 (1964).

[26] S. M. Sze, Physics of Semiconductor Devices (Wiley, New York, 1981).

[27] L. J. Brillson, Phys. Rev. Letters 40, 260 (1978), and references therein.

[28] R. H. Williams, V. Montgomery, and R. R. Varma, J. Phys. C 11, L735 (1978).

[29] R. H. Williams and M. H. Patterson, Appl. Phys. Lett. 40, 484 (1982).

[30] R. H. Williams, to be published.

[31] M. S. Daw and D. L. Smith, Phys. Rev. B20, 5150 (1979); J. Vac. Sci. Technol. 17, 1028 (1980); Appl. Phys. Lett. 36, 690 (1980); Solid State Commun. 37, 205 (1981); M. S. Daw, D. L. Smith, C. A. Swarts, and T. C. McGill, J. Vac. Sci. Technol. 19, 508 (1981).

[32] R. E. Allen and J. D. Dow, Phys. Rev. B24, 911 (1981).

[33] R. E. Allen and J. D. Dow, J. Vac. Sci. Technol. 19, 383 (1981).

[34] R. E. Allen and J. D. Dow, Appl. Surf. Sci. 11/12, 362 (1982).

[35] R. E. Allen, J. D. Dow, and H. P. Hjalmarson, Solid State Commun. 41, 419 (1982).

[36] J. D. Dow and R. E. Allen, J. Vac. Sci. Technol. 20, 659 (1982).

[37] R. E. Allen, R. P. Beres, and J. D. Dow, J. Vac. Sci. Technol. B1, 401 (1983).

[38] H. H. Wieder, Inst. Phys. Conf. Ser. 50, 234 (1980).

[39] H. H. Wieder, Appl. Phys. Lett. 38, 170 (1981).

[40] H. H. Wieder, to be published.

[41] W. Mönch and H. J. Clemens, J. Vac. Sci. Technol. 16, 1238 (1979).

[42] H. Gant and W. Mönch, Appl. Surf. Sci. 11/12, 332 (1982).

[43] W. Mönch and H. Gant, Phys. Rev. Letters 48, 512 (1982).

[44] W. Mönch, R. S. Bauer, H. Gant, and R. Murschall, J. Vac. Sci. Technol. 21, 498 (1982). J. Assmann and W. Mönch, Surf. Sci. 99, 34 (1980).

[45] W. Mönch, to be published.

[46] O. F. Sankey, R. E. Allen, and J. D. Dow, Solid State Commun. 49, 1 (1983).

[47] J. L. Freeouf and J. M. Woodall, Appl. Phys. Lett. 39, 727 (1981); J. L. Freeouf, Solid State Commun. 30, 1059 (1980); J. M. Woodall, G. D. Petit, T. N. Jackson, C. Lanza, K. L. Kavanagh, and J. W. Mayer, Phys. Rev. Letters 51, 1783 (1983).

[48] R. Ludecke and L. Esaki, Phys. Rev. Letters 33, 653 (1974); R. Ludecke and A. Kona, Phys. Rev. Letters 34, 817 (1975); Phys. Rev. Letters 39, 1042 (1977); and to be published.

[49] P. S. Ho and G. W. Rubloff, Thin Solid Films, 89, 433 (1981).

[50] J. M. Andrews and J. C. Phillips, Phys. Rev. Letters 35, 56 (1975).

[51] A. A. Maradudin, E. W. Montroll, and G. H. Weiss, Solid State Phys. Suppl. 3, 132 (1963).

[52] M. Lannoo and P. Lenglart, J. Phys. Cs 30, 2409 (1969).

[53] S. T. Tong, A. R. Lubinsky, B. J. Mrstik, and M. A. van Hove, Phys. Rev. B17, 3303 (1978).

[54] O. F. Sankey and J. D. Dow, Phys. Rev. B27, 7641 (1983).

[55] O. F. Sankey, H. P. Hjalmarson, J. D. Dow, D. J. Wolford, and B. G. Streetman, Phys. Rev. Letters 45, 1656 (1980); O. F. Sankey and J. D. Dow, Appl. Phys. Lett. 38, 685 (1981); J. Appl. Phys. 52, 5139 (1981); Phys. Rev. B26, 3243 (1982).

[56] H. Wieder, to be published.

[57] H. H. Wieder, private communication.

[58] W. Mönch, to be published.

[59] P. Vogl, H. P. Hjalmarson, and J. D. Dow, J. Phys. Chem. Solids 44, 365 (1983).

[60] W. M. Gibson, to be published, has recently proposed that the relaxation of the (110) surface of GaAs is considerably less than once thought [53], based on his He-atom scattering data.

[61] A. Kahn, E. So, P. Mark, and C. B. Duke, J. Vac. Sci. Technol. 15, 580 (1978).

[62] C. B. Duke, R. J. Meyer, and P. Mark, J. Vac. Sci. Technol. 17, 971 (1980).

[63] C. B. Duke, A. Paton, W. K. Ford, A. Kahn, and J. Carelli, Phys. Rev. B24, 562 (1981); C. B. Duke, A. Paton, W. K. Ford, A. Kahn, and G. Scott, Phys. Rev. B24, 3310 (1981).

[64] J. A. Knapp and G. J. Lapeyre, J. Vac. Sci. Technol. 13, 757 (1976); Nuovo Cimento 39B, 693 (1977).

[65] G. P. Williams, R. J. Smith, and G. J. Lapeyre, J. Vac. Sci. Technol. 15, 1249 (1978).

[66] A. Huijser, J. van Laar, and T. L. Rooy, Phys. Lett. 65A, 337 (1978).

[67] J. van Laar and J. J. Scheer, Surf. Sci. 8, 342 (1967); J. van Laar and A.

Huijser, J. Vac. Sci. Technol. 13, 769 (1976).

[68] W. Gudat and D. E. Eastman, J. Vac. Sci. Technol. 13, 831 (1976).

[69] A. Huijser and J. van Laar, Surf. Sci. 52, 202 (1975); A. Huijser, J. van Laar, and T. L. van Rooy, Surf. Sci. 62, 472 (1977).

[70] G. M Guichar, C. A. Sebenne, and C. D. Thualt, J. Vac. Sci. Technol. 16, 1212 (1979); D. Norman, I. T. McGovern, and C. Norris, Phys. Lett. 63A, 384 (1977).

[71] P. Chiaradia, G. Chiarotti, F. Ciccacci, R. Memeo, S. Nannarone, P. Sassaroli, and S. Selci, Surf. Sci. 99, 76 (1980).

[72] V. Dose, H.-J. Gassmann, and D. Straub, Phys. Rev. Letters 47, 608 (1981).

[73] Further work is cited in the papers above and in the review by J. Pollmann, Festkörperprobleme, Vol. XX, p. 117.

[74] A. Zunger, Phys. Rev. B22, 959 (1980).

[75] See the proceedings of the Ninth International Conference on the Physics and Chemistry of Semiconductor Interfaces; J. Vac. Sci. Technol. 21 (1982).

[76] D. Haneman, Phys. Rev. 121, 1093 (1961).

[77] R. Feder, W. Mönch, and P. P. Aver, J. Phys. C 12, 2179 (1979).

[78] J. P. Buisson, J. D. Dow, and R. E. Allen, Surf. Sci. 120, L477 (1982), and references therein.

[79] K. C. Pandey, Phys. Rev. Letters 47, 1913 (1981).

[80] D. E. Eastman, J. Vac. Sci. Technol. 17, 492 (1980).

[81] D. J. Chadi, Appl. Optics 19, 3971 (1980).

[82] A. Redondo and W. A. Goddard, J. Vac. Sci. Technol. 21, 344 (1982).

[83] D. J. Chadi, Phys. Rev. Letters 43, 43 (1979).

[84] J. Ihm, M. L. Cohen, and D. J. Chadi, Phys. Rev. B21, 4952 (1980).

[85] M. A. Bowen, J. D. Dow, and R. E. Allen, Phys. Rev. B26, 7083 (1982).

[86] A. Mazur, J. Pollmann, and M. Schmeits, Phys. Rev. B26, 7086 (1982).

[87] W. Mönch, P. Koke, and S. Krueger, J. Vac. Sci. Technol. 19, 313 (1981).

[88] R. E. Allen and J. D. Dow, Phys. Rev. B24, 911 (1981). See also, R. E. Allen, H. P. Hjalmarson, H. Büttner, P. Vogl, D. J. Wolford, O. F. Sankey, and J. D. Dow, Internatl. J. Quantum Chem., Quantum Chem. Sympos. 14, 607 (1980).

[89] R. P. Beres, R. E. Allen, and J. D. Dow, Phys. Rev. B26, 769 (1982).

[90] T. Takahashi and A. Ebina, Appl. Surf. Sci. 11/12, 268 (1982).

FIGURE CAPTIONS

Fig. 1. Schematic illustration of Fermi-level pinning. Band edges of the bulk semiconductor, the semiconductor surface, and the Fermi energy of the metal, the surface of the semiconductor, and the semiconductor are all shown as functions of position. The lowest energy surface defect level that is not fully occupied (before charge is allowed to flow) is denoted by an open circle. This level and the Fermi levels of the n-type semiconductor and the metal align.

Fig. 2. Surface Fermi energy of n-type InP versus heat of reaction of InP with the metals Ni, Fe, Al, Cu, Ag, and Au, extracted from data of Ref. [28], assuming Fermi-level pinning. The theoretical Fermi-level pinning defect levels for the surface P-vacancy (V_P), the native antisite defects (In_P and P_{In}), and the extrinsic impurities S on a P-site (S_P) and Sn on a surface In site (Sn_{In}) are given at the right of the figure. The n-InP data can be interpreted as follows: non-reactive metals produce only antisite defects as the dominant defects; reactive metals and treatment of the surface with oxygen and Cl produce P-vacancies. Treatments with Sn and S produce surface Sn_{In} and S_P as dominat defects, respectively.

Fig. 3. The totally symmetric (a_1) levels for a bulk Si vacancy, surrounded by one Si atom and three X atoms, as a function of the defect potential V, normalized to the Ni defect potential, after Ref. [46]. For V=0, the X atoms are Si; for $V=V_{Ni}$, the X atoms are Ni. The parent levels of the isolated Si vacancy are shown for V=0. The experimental Fermi-level pinning position for $NiSi_2$ extracted from the data of G. Ottavianai, K. N. Tu, and J. W. Mayer, Phys. Rev. B24, 3354 (1981) are denoted by a dot with a label $NiSi_2$.

Fig. 4. One type of interfacial vacancy "sheltering" a Si dangling bond, after Ref. [46]. The geometry is that determined for the $NiSi_2/Si(111)$ interface determined by D. Cherns, G. R. Anstis, J. L. Hutchison, and J. C. H. Spence, Phil. Mag. A46, 849 (1982).

Fig. 5. Predicted dependence of Schottky barrier height on alloy compositions x and y of $In_{1-x}Ga_xAs$ and $Ga_{1-y}Al_yAs$ alloys, compared with data, after Ref. [88].

Fig. 6. Predicted surface state dispersion curves $E(\bar{k})$ for surface bound states (solid lines) and surface resonances (dashed lines) at the relaxed (110) surface of GaAs, after Ref. [14]. The energy is plotted as a function of the planar wavevector $\bar{k}$ along the symmetry lines of the surface Brillouin zone, shown on the right. The labelling is the same as that of Chelikowsky and Cohen (Ref. [7]), with A_1 A_2, C_1, and C_2 mainly s-like, and A_3 A_4, A_5, C_3, and C_4 mainly p-like. A_5 and C_3 are the "dangling-bond" states. A_3, A_1', and A_2' are largely associated with in-plane p-orbitals in the first and second layers. The character of each state varies somewhat with the planar wavevector $\bar{k}$, and represents an admixture of all orbitals. The widths of the resonances are typically 0.5 to 1.0 eV, but in some cases are smaller than 0.1 eV or as large as 2.0 to 5.0 eV. The dots follow the continuous disperson curves inferred by Huijser et al. (Ref. [66]) for the "clear" and "weak" experimental features. The open squares represent the states observed by Williams et al. (Ref. [65]). The data reported in Refs. [64] and [6] are consistent with those shown here.

Fig. 7. Predicted energies of surface bound states (solid lines) and surface resonances (dashed) for the (110) surface of ZnSe, as function of the planar wavevector $\bar{k}=(k_1,k_2)$, after Ref. [15]. The surface Brillouin zone is shown on the right; $\bar{\Gamma}$ is the origin, $\bar{k}=(0,0)$. The bulk bands are shaded. E_v and E_c are the valence and conduction band edges. The experimental features identified with bound and resonant surface states in Ref. [11], along the two symmetry lines $\bar{\Gamma}\bar{X}'$ and $\bar{X}\bar{\Gamma}$, are indicated by the dotted lines.

Fig. 8. Dispersion curves for surface states and surface resonances at the (100) (2×1) surface of Si, after Ref. [85]. The energy E is shown as a function of the planar wavevector $\bar{k}$ around the symmetry lines of the surface Brillouin zone. Solid lines represent results of the present calculations; dashed lines are the measurements of R. I. G. Uhrberg, G. V. Hansson, J. M. Nicolle, and S. A. Flodstrom, Phys. Rev. B24, 4684 (1981); and the dotted line is the measurement of F. J. Himpsel and D. E. Eastman, J. Vac. Sci. Technol. 16, 1297 (1979), which were taken from $\bar{\Gamma}$ to $\bar{J}'$ along the (010) direction, rather than along the symmetry line $\bar{\Gamma}$ to $\bar{J}'$. E_v and E_c are the Si valence and conduction band edges.

Fig. 9. Predicted and observed Ga 3d core surface Frenkel excitons (double lobes) for GaAs, GaSb, and GaP, after Ref. [32]. The lower unoccupied surface states (Ref. [13][14]) are represented by closely spaced horizontal lines. E_v and E_c are, respectively, the top of the valence band and the bottom of the conduction band. The experimental results here and in Fig. 7 are those of Eastman and co-workers (D. E. Eastman and J. L. Freeouf, prl 33, 1601 (1974); 34, 1624 (1975); W. Gudat and D. E. Eastman, J. Vac. Sci. Technol. 13, 831 (1976); D. E. Eastman, T.-C. Chiang, P. Heimann, anf F. J. Himpsel, prl

45, 656 (1980).).

Fig. 10. Predicted and observed In 4d core surface Frenkel excitons for InAs, InSb, and InP, after Ref. [32].

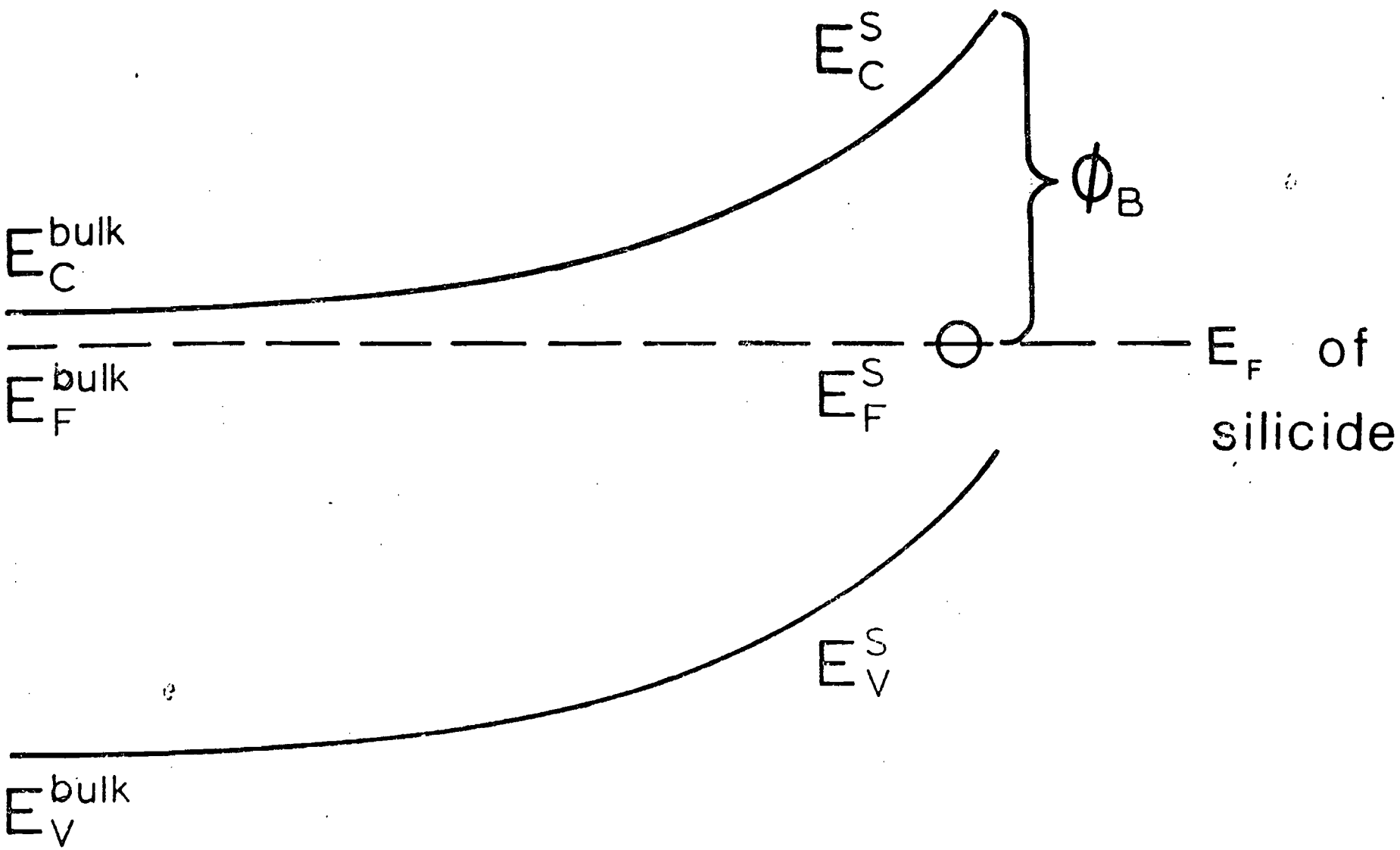

E_C^S
ϕ_B
E_C^{bulk}
E_F^{bulk}
E_F^S
E_F of silicide
E_V^S
E_V^{bulk}
Si
Interface

Fig. 1

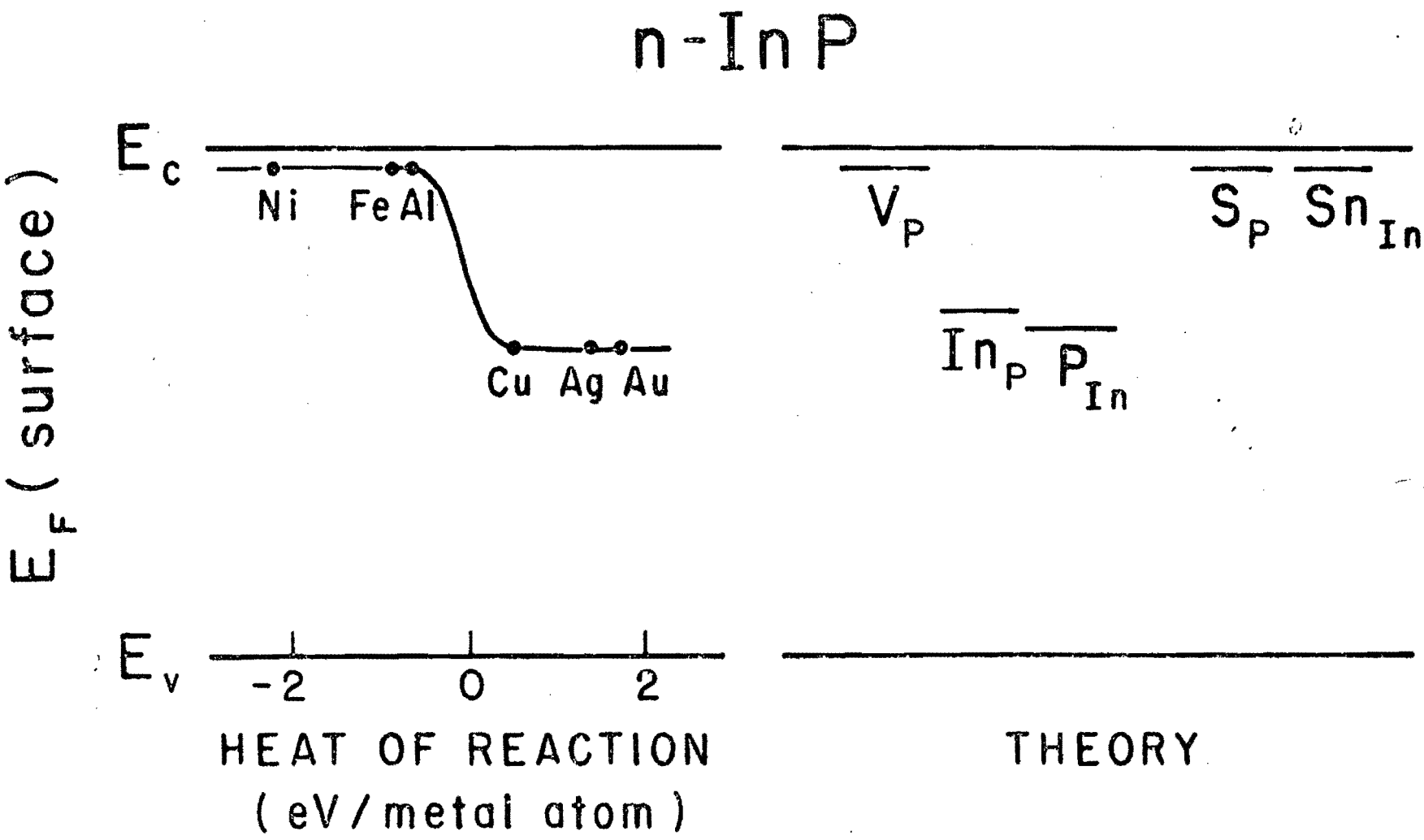

n-InP
E_c
Ni
Fe Al
Cu Ag Au
$\overline{V_P}$
$\overline{S_P}$ $\overline{Sn_{In}}$
$\overline{In_P}$ $\overline{P_{In}}$
E_v
-2
0
2
E_F (surface)
HEAT OF REACTION
(eV / metal atom)
THEORY

Fig. 2

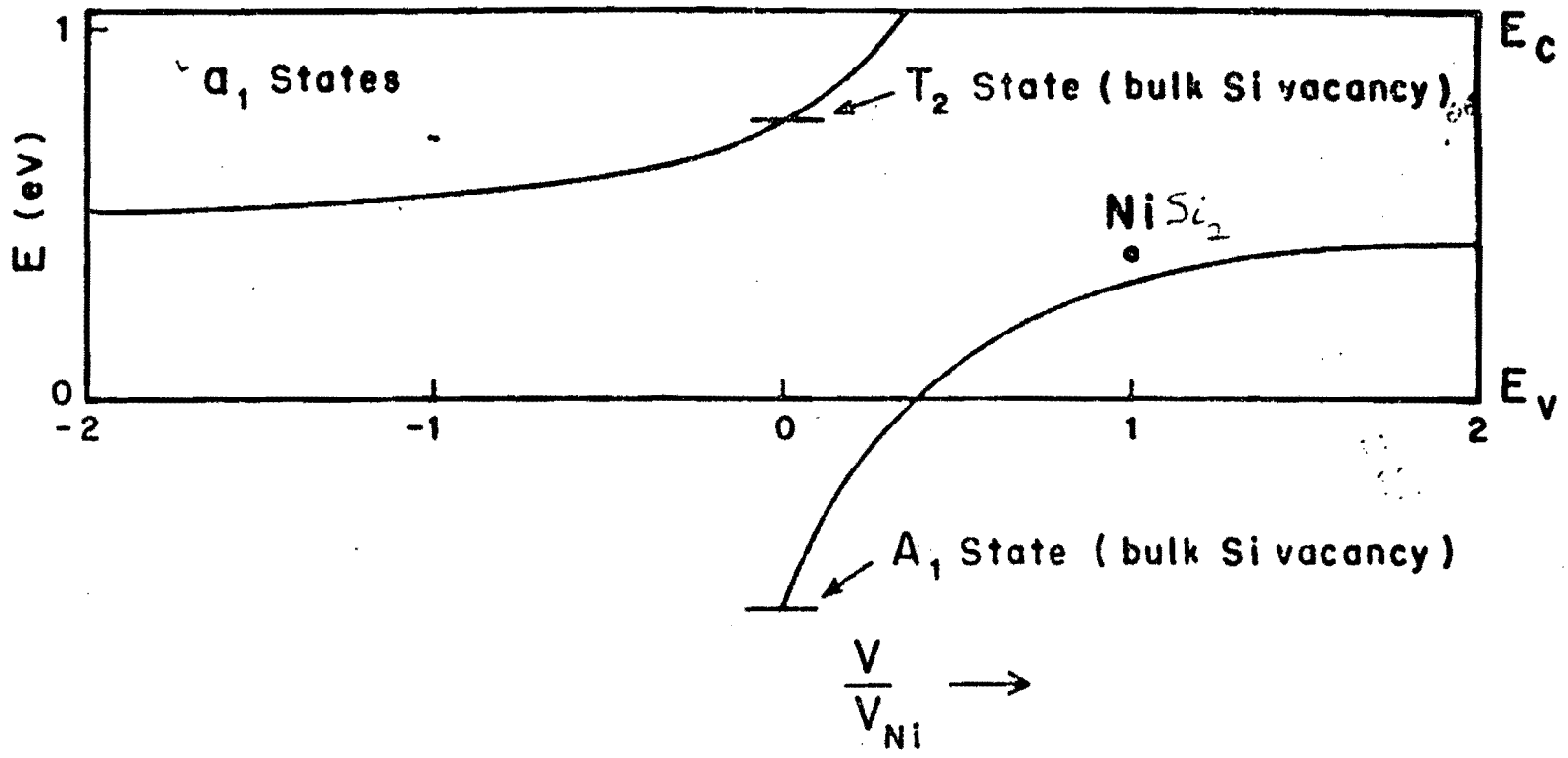

1
0
E (eV)
-2
-1
0
1
2
E_C
E_V
a_1 States
T_2 State (bulk Si vacancy)
$NiSi_2$
A_1 State (bulk Si vacancy)
$\frac{V}{V_{Ni}}$ →

Fig. 3

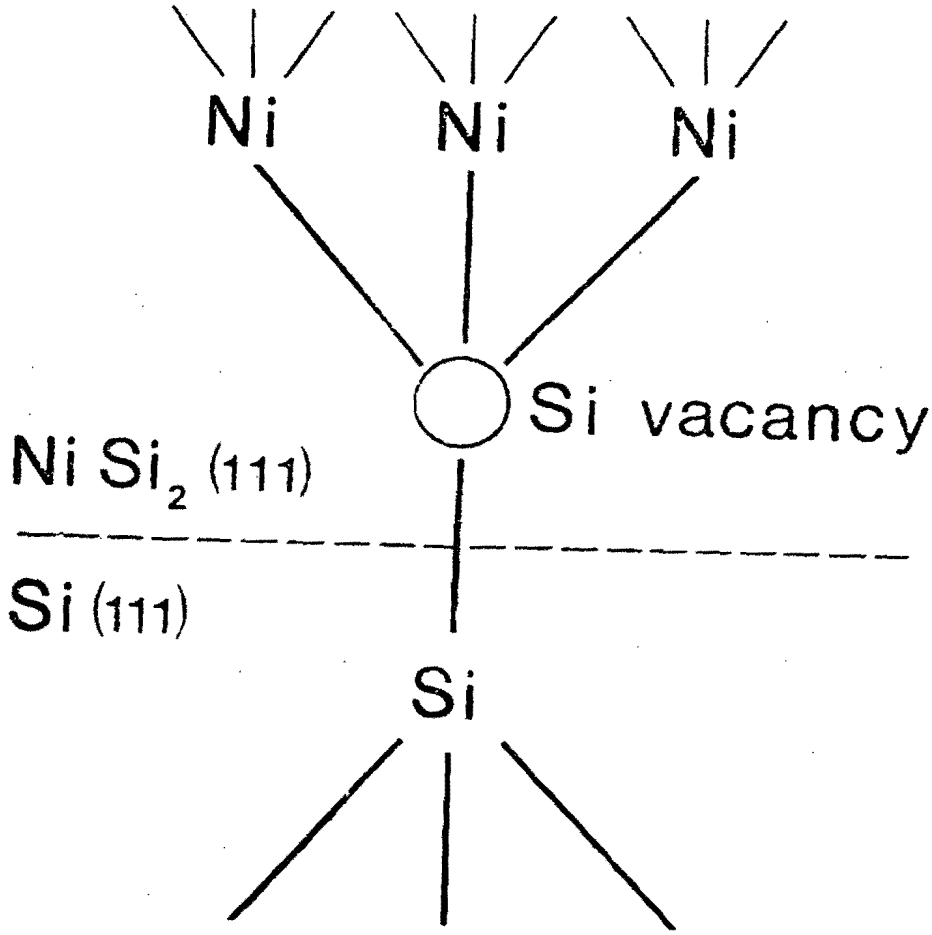


Fig. 4

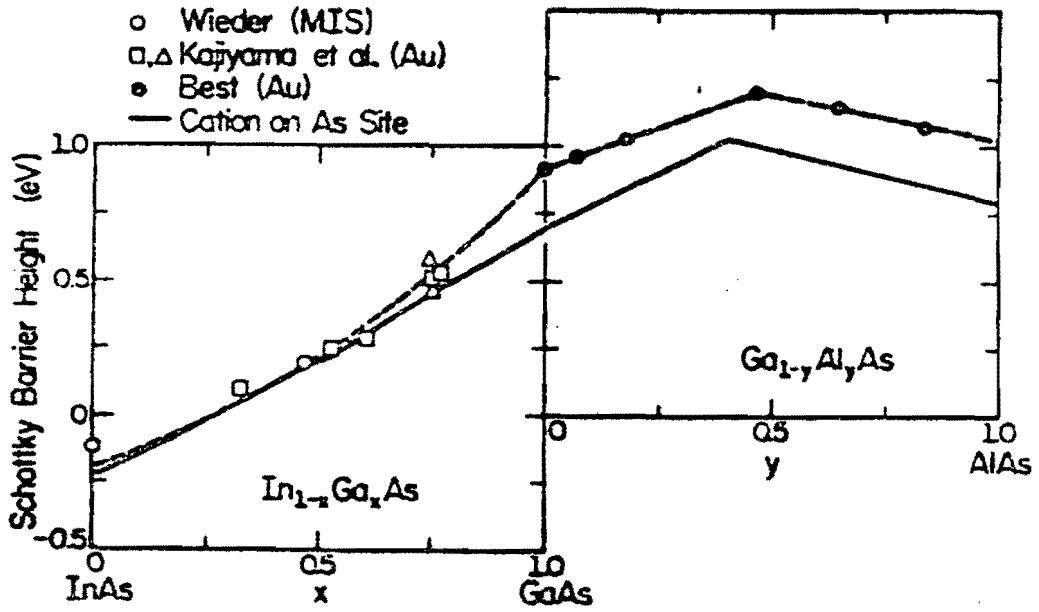

o Wieder (MIS)
□,△ Kajiyama et al. (Au)
• Best (Au)
— Cation on As Site
Schottky Barrier Height (eV)
1.0
0.5
0
-0.5
$In_{1-x}Ga_xAs$
0
InAs
0.5
x
1.0
GaAs
$Ga_{1-y}Al_yAs$
0
0.5
y
1.0
AlAs

Fig. 5

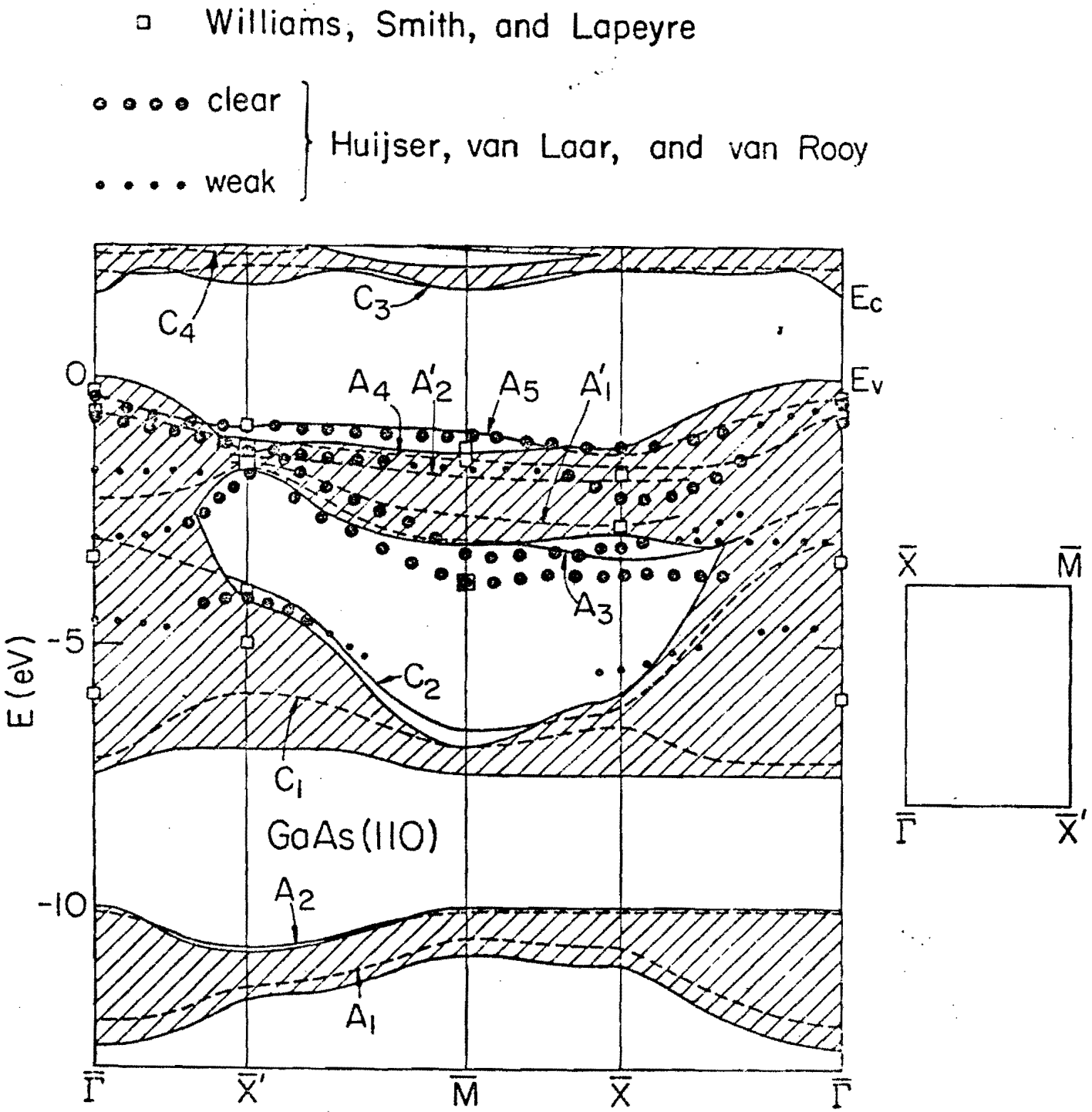

Williams, Smith, and Lapeyre
clear
weak
Huijser, van Laar, and van Rooy
C4
C3
Ec
0
A4
A'2
A5
A'1
Ev
A3
-5
E (eV)
C2
C1
GaAs(110)
-10
A2
A1
Γ̄
X̄'
M̄
X̄
Γ̄
X̄
M̄
Γ̄
X̄'

Fig. 6

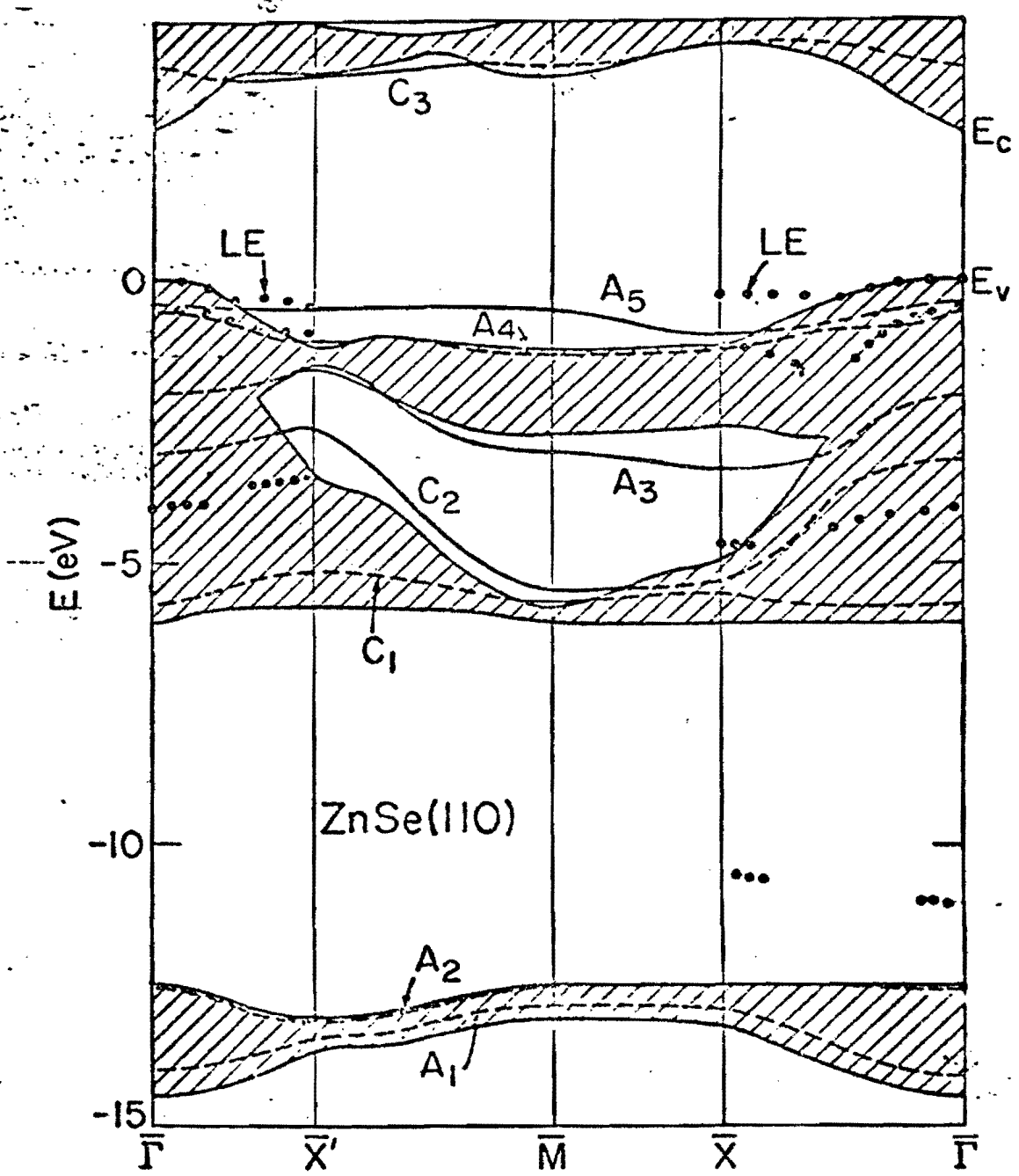

C3
Ec
LE
LE
0
A5
Ev
A4,
A3
C2
E (eV)
-5
C1
ZnSe(110)
-10
A2
A1
-15
Γ̄
X̄'
M̄
X̄
Γ̄
X̄
M̄
Γ̄
X̄'
Experiment •••••
(Γ̄X̄' and X̄Γ̄ only)

Fig. 7

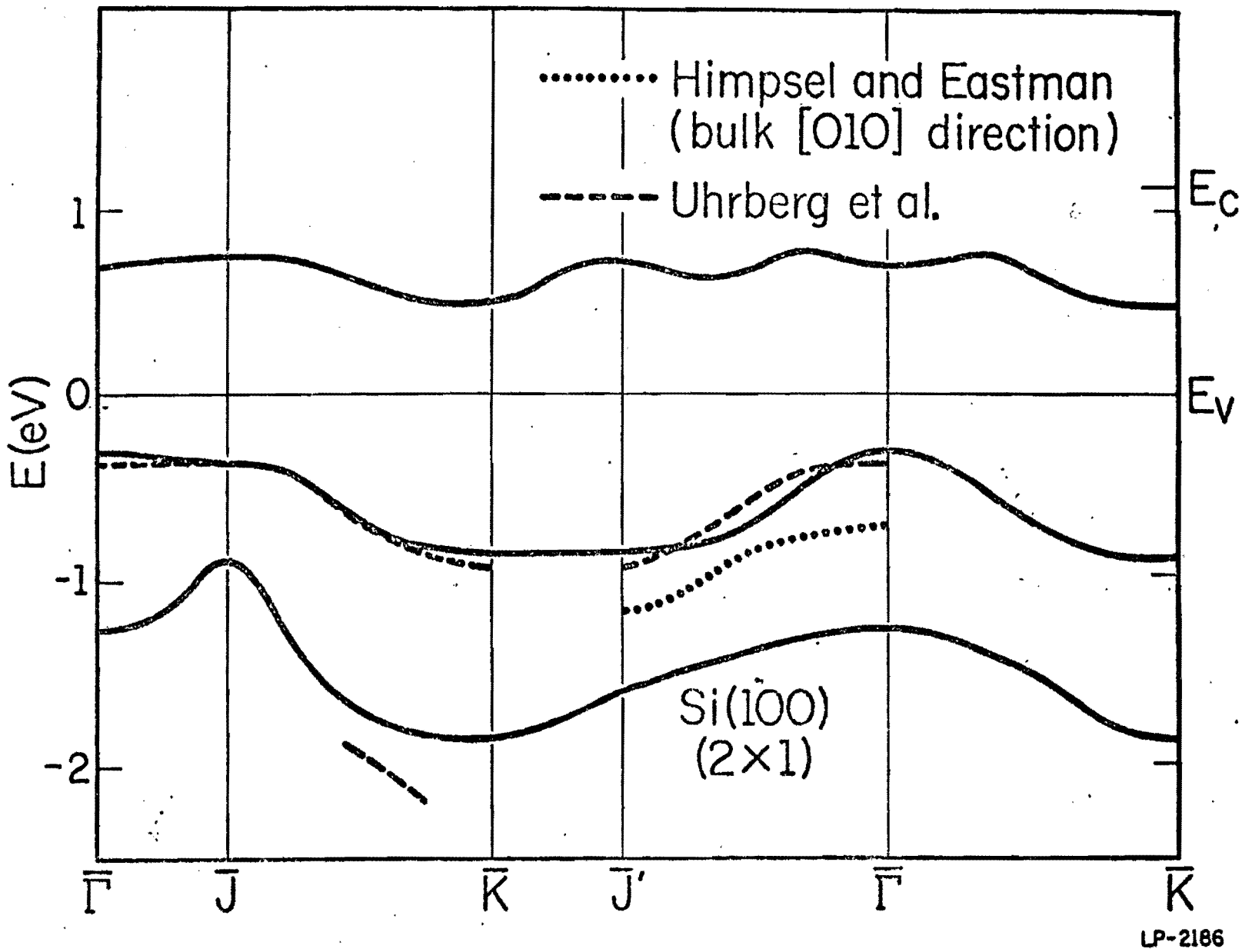
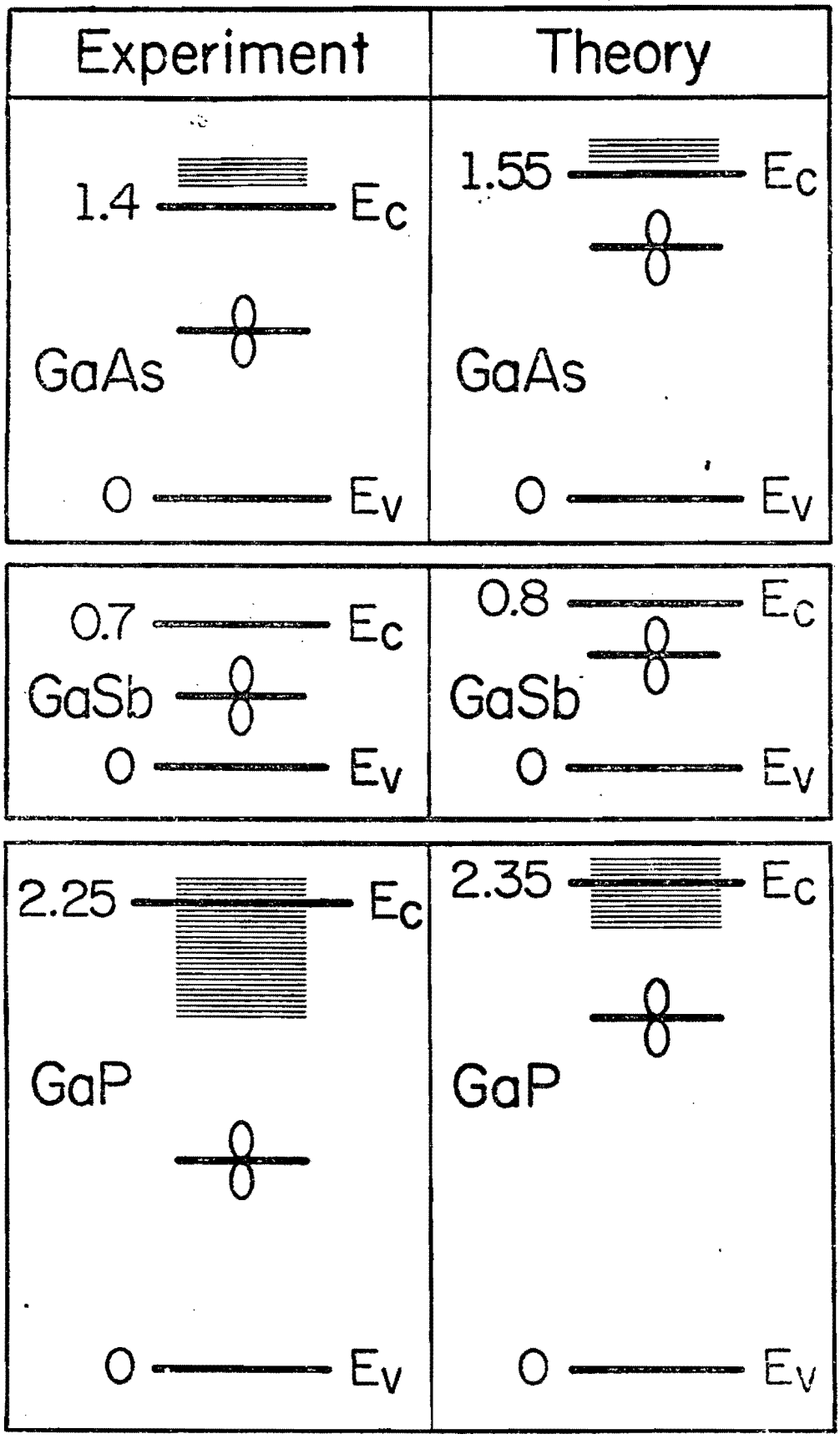
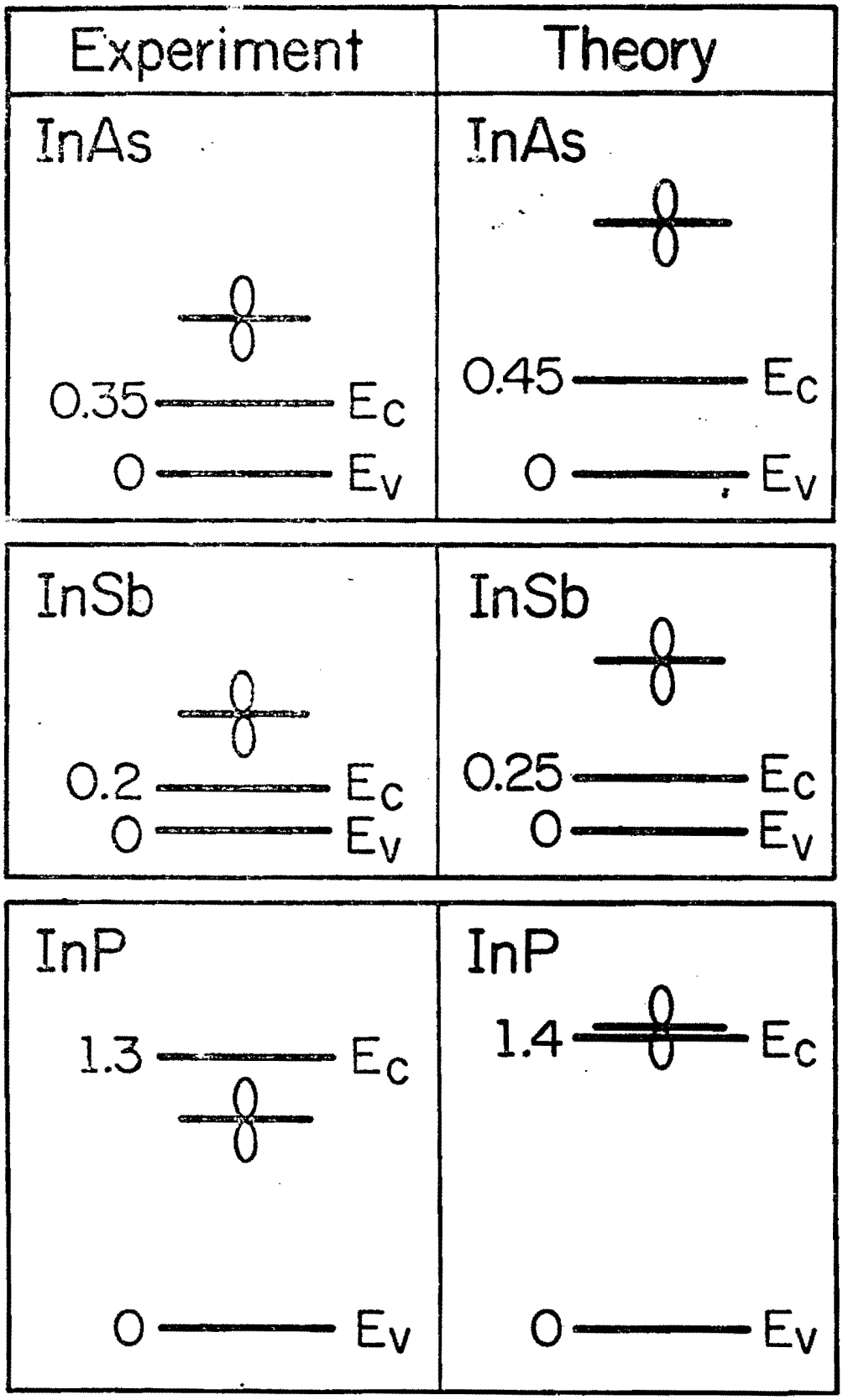

Himpsel and Eastman
(bulk [010] direction)
Uhrberg et al.
E_c
E_v
E(eV)
1
0
-1
-2
Si(100)
(2×1)
$\bar{\Gamma}$
$\bar{J}$
$\bar{K}$
$\bar{J}'$
$\bar{\Gamma}$
$\bar{K}$
LP-2186

Fig. 8


Experiment
Theory
1.4
Ec
GaAs
0
Ev
1.55
Ec
GaAs
0
Ev
0.7
Ec
GaSb
0
Ev
0.8
Ec
GaSb
0
Ev
2.25
Ec
GaP
0
Ev
2.35
Ec
GaP
0
Ev

LP-1844

Fig. 9



LP-1846

Fig. 10

Ultramicroscopy 14 (1984) 127–130
North-Holland, Amsterdam

THEORY OF Si/TRANSITION-METAL SILICIDE SCHOTTKY BARRIERS

Otto F. SANKEY

Department of Physics, Arizona State University, Tempe, Arizona 85287, USA

Roland E. ALLEN

Department of Physics, Texas A&M University, College Station, Texas 77843, USA

and

John D. DOW

Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556, USA

Received 20 January 1984; presented at Workshop January 1984

The height of the Schottky barrier formed at transition-metal/Si interfaces varies over a very small range (~ 0.4 eV) considering the wide range of electronic structures possible from one end of the transition-metal series to the other. Furthermore, the barriers are observed to form within a few monolayers of metal coverage, suggesting that the barrier is a property of the local bonding and that the true metallic states play only a minor role. A model has been developed to explain these facts in terms of the Fermi-level pinning mechanism of Schottky barrier formation. The physics contained in the model is that of a Si dangling bond sheltered from the transition-metal-silicide by an interfacial vacancy. Since (i) the dangling-bond is sheltered from the metallic-silicide and (ii) the atomic energy levels of the transition metal are out of resonance with Si, the dangling bond (which forms a level in the Si band gap) will be only weakly perturbed by the silicide. Thus this interfacial dangling bond can pin the Fermi level at nearly the same energy for all the transition-metal-silicides. A tight-binding calculation of the electronic structure of this defect at the $NiSi_2/Si(111)$ interface has been performed for an infinite interface using the transfer-matrix technique. The results of this calculation are described in terms of a very simple molecular model.

It is a remarkable fact that the Schottky barrier heights for the whole range of Si/silicide interfaces varies over a relatively narrow range of about 0.55–0.87 eV in n-Si [1]. For a Si bandgap of 1.1 eV, this places the Fermi-level in the *lower* part of the bandgap between 0.23–0.55 eV above the valence band edge. Here we argue that such barriers can be understood in terms of Fermi-level pinning [2] by a small concentration of Si dangling bonds that are "sheltered" from the transition metal by vacancies at the Si/silicide interface. This explanation, which differs substantially from previous theories of Si Schottky barrier formation [1,3,4], unifies the understanding of Si/transition-metal Schottky barriers with the generally accepted model of Fermi-level pinning by native defects [5,6] at (110) interfaces between III–V semiconductors and metals [7] or other overlayers. A more detailed account of our work will be given elsewhere [8].

The following observations place severe constraints on any theory of Schottky barrier formation at Si/silicide interfaces: (1) The barrier heights for the silicides all lie within 0.4 eV of one another for all the different transition metals, stoichiometries, and crystal structures. (2) The barriers are observed to form at low coverages before a complete metallic silicide is formed, indicating that the local atomic bonding at the interface, rather than any collective interface property, determines the barrier [1]. (3) There are only slight variations of the barrier height for different compounds of a

0304-3991/84/$03.00

given transition metal [9,10] indicating that the transition metal itself, rather than stoichiometry, crystal structure, etc., determines the barrier. (4) The barrier heights for n- and p-type Si very nearly add up to the Si bandgap. This means that the pinning level must only be partially occupied, so that it may act as both an acceptor and a donor.

We propose that these observations can be understood quite naturally in terms of an interfacial vacancy which shelters a Si dangling bond from the effects of the transition metal. This dangling bond has only a weak link with the silicide and is only slightly perturbed by the transtition metal s-, p-, and d-orbitals, and hence is insensitive to the large variations (on a 1eV scale) one might expect to occur when the transition metal is varied or when the stoichiometry or the crystal structure of the silicide is changed.

To make these ideas specific, we consider a particular example of such a defect – the example illustrated in fig. 1 for the case of the abrupt $Si/NiSi_2(111)$ interface. If the vacancy in fig. 1 were replaced by a Si atom, one would have the bonding configuration determined by Cherns et al. [11]. For the reactive systems under consideration here (transition metals "eating" their way into Si), a reasonable concentration of vacancies ($\sim 10^{13}$ cm^{-2}) appears quite likely.

Here we consider a very simple model of the electronic structure of the defect shown in fig. 1. This model is justified only by the results of the more complete calculation described elsewhere [8], but it reveals the essential physics of the problem.

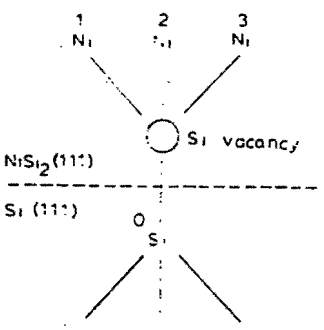


Fig. 1. An example of an interfacial vacancy sheltering a Si dangling bond. Replacing the vacancy by a Si atom gives the geometry of the $NiSi_2/Si$ (111) interface determined by Cherns et al. [11].

In this simplified model, only four atoms are explicitly considered – those surrounding the vacancy – and only one sp^3 hybrid orbital per atom – which is directed toward the vacancy. We first take all four atoms to be Si (tetrahedral symmetry) and later change three of these atoms into Ni (C_{3v} symmetry) to simulate the $Si/NiSi_2$ interface.

Taking all four atoms to be Si, we construct the A_1 and T_2 states of the bulk Si vacancy:

$$|A_1(a_1)\rangle = \tfrac{1}{2}(|\phi_0\rangle + |\phi_1\rangle + |\phi_2\rangle + |\phi_3\rangle), \quad (1a)$$

$$|T_2(a_1)\rangle = (1/\sqrt{12})(3|\phi_0\rangle - |\phi_1\rangle - |\phi_2\rangle - |\phi_3\rangle), \quad (1b)$$

$$|T_2(e),1\rangle = (1/\sqrt{2})(|\phi_1\rangle - |\phi_3\rangle), \quad (1c)$$

$$|T_2(e),2\rangle = (1/\sqrt{6})(|\phi_1\rangle + |\phi_3\rangle - 2|\phi_2\rangle), \quad (1d)$$

where the orbital $|\phi_i\rangle$ is the hybrid orbital of atom i. The energies of the A_1 and T_2 levels can be described by two parameters ϵ_h and t; here we have $\epsilon_h = \langle\phi_i|H|\phi_i\rangle$ is the orbital energy of an sp^3 hybrid, and $-t = \langle\phi_i|H|\phi_j\rangle$ for $i \neq j$ represents the interaction between two different hybrid orbitals. These two parameters represent effective interactions and are obtained by fitting to the bulk Si vacancy deep levels. The A_1 level is resonant with the valence band at $E_{A_1} = \epsilon_h - 3t$, while the triply degenerate T_2 level lies in the Si bandgap and has an energy $E_{T_2} = \epsilon_h + t$. We list in table 1

Table 1
The calculated A_1- and T_2-symmetric energy levels for the unrelaxed Si vacancy by several workers; the calculations are either pseudopotential (P) or tight-binding (TB); all energies are in eV, and the top of the valence band is defined to be the zero of energy; the Si bandgap is 1.1eV; the two parameters ϵ_h and t are simply obtained from the A_1 and T_2 energy levels (see text); the important parameter ϵ_h is the energy of a single dangling bond and is found to lie in the lower part of the Si bandgap in all cases

Type of calculation		A_1 level	T_2 level	t	ϵ_h
P	TB				
Ref. [12]		−1.10	0.70	0.45	0.25
Ref. [13]		−0.60	0.80	0.35	0.45
Ref. [14]		−1.10	0.60	0.42	0.16
	Ref. [15]	−0.55	0.75	0.33	0.43
	Ref. [16]	−0.96	0.51	0.37	0.14

the A_1 and T_2 levels for three different pseudopotential calculations [12–14] and two tight-binding [15,16] calculations for the unrelaxed vacancy in Si. From these levels the parameters ϵ_h and t can be extracted using $\epsilon_h = (E_{A_1} + 3E_{T_2})/4$ and $t = (E_{T_2} - E_{A_1})/4$. Note that ϵ_h, which is the energy of a *single* Si dangling bond, lies in the lower part of the Si bandgap in all cases.

We next change three of the atoms surrounding the vacancy into Ni atoms. We do this by raising the hybrid orbital energies of atoms 1, 2, and 3 (see fig. 1) from ϵ_h to $\epsilon_h + V$, where $V \approx 5$ eV [8] represents the (large) positive difference between a Ni and a Si sp^3 hybrid orbital. The symmetry is now reduced from T_d to C_{3v}, and the possible levels are of a_1 (σ-like) and e (π-like) symmetry. The states of e-symmetry evolve from two of the T_2 levels of the bulk Si vacancy (the T_2(e) levels in eqs. (1c) and (1d)), but are raised out of the gap roughly linearly with the potential V to become resonant with the conduction bands. Since the e-symmetric levels are not in the gap and are metal-atom derived, they play no role in pinning the interfacial Fermi level, and we will no longer consider them.

The interesting levels are those of a_1-symmetry which are admixtures of the $|A_1(a_1)\rangle$ (eq. (1a)) and the $|T_2(a_1)\rangle$ (eq. (1b)) levels of the bulk Si vacancy. However, since the Si and Ni hybrid orbitals are no longer degenerate, perturbation theory shows that the effective interaction between Si and Ni hybrid orbitals is reduced from t (~ 0.4 eV) for the bulk Si vacancy to t^2/V (~ 0.03 eV) for the interfacial vacancy. A schematic energy level diagram for a_1-symmetric states of the bulk and interfacial vacancies is shown in fig. 2. Note that because Ni (or any transition-metal element) and Si are "out of resonance", a level is formed in the lower part of the Si bandgap which is tied to the Si dangling bond energy ϵ_h and is relatively insensitive to the transition metal as long as we have $V \gg t$. This simple model leads to the important conclusion that for various transition metals, interfacial Fermi-level pinning positions are nearly equal to, but slightly below, the "defect pinning energy" ϵ_h of a single Si dangling bond.

We briefly mention the more rigorous calculations on which the simple model is based. A

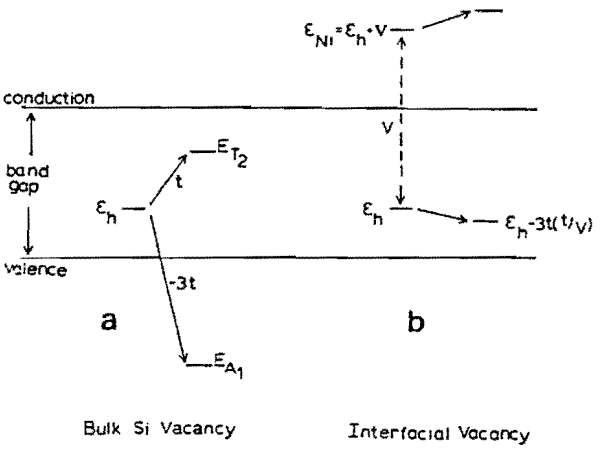


Fig. 2. Schematic energy level diagram of the a_1-symmetric levels of (a) the bulk Si vacancy and (b) the interfacial vacancy. In (a) the hybrid orbitals at ϵ_h lie in the lower part of the Si bandgap but interact strongly through t to produce the A_1 level resonant with the valence band and a T_2 level in the upper part of the bandgap. In (b) the hybrid orbitals of Si and Ni are no longer degenerate and their interaction is reduced by t/V. This brings E_{A_1} out of the valence band so that it now lies only slightly below the Si dangling bond energy ϵ_h. (The a_1-symmetric representation refers to the C_{3v} group appropriate for the interfacial vacancy. Since C_{3v} is a subgroup of T_d, the A_1 and one of the T_2 levels of the bulk Si vacancy are also a_1-symmetric.)

tight-binding calculation was performed for an embedded cluster of a vacancy and three Ni atoms (including d-orbitals on Ni) in an infinite Si host [8] using the Si tight-binding model of ref. [15]. The Si dangling-bond-like level is found at 0.4 eV. The d-orbitals are found to play only a minor role. Since the d-orbital energies lie well below the Fermi level, they tend to push up slightly on the Si dangling bond, but with a greatly reduced strength because the d-orbital is not a nearest neighbor to the dangling bond orbital and hence interacts with it either through a small second-neighbor interaction or indirectly via its interactions with the intervening Si atoms surrounding the vacancy. (The d-orbitals were taken to interact only with nearest neighbors.) More sophisticated calculations for an interface between semi-infinite slabs of $NiSi_2$ and Si have recently been completed [17] using the transfer-matrix technique [18]. The tight-binding bands of $NiSi_2$ have been fit to the bulk bands of Chabal et al. [19], and the tight-binding model of

Vogl et al. [20] has been used for bulk Si. Here the interfacial vacancy level is found to lie at 0.13 eV. Although the two calculations give slightly different results, and the estimates made for the dangling bond energy ϵ_h in table 1 differ by ~ 0.3 eV, they all show that the defect "pinning" level lies in the lower part of the Si bandgap. Measurements for a Si dangling bond quite similar to the one described here [21] at the SiO_2/Si interface show a level at 0.36 eV [22]. As mentioned earlier, the interfacial Fermi level for the silicides lie approximately in the range 0.23–0.55 eV.

Conclusion. The present theory is manifestly based on local atomic bonding and a localized defect, and is thus compatible with the experimental findings [1,9,10] that the observed Schottky barriers form before the completion of a complete metallic overlayer. Since in this model the barriers are determined mainly by Si, the barrier is affected to a lesser degree by the nature of the transition metal atom, stoichiometry, or crystal structure of the silicide. Furthermore, since the dangling bond is occupied by a single electron, it can act either as a donor or an acceptor – this leads to very nearly the same pinning position for both n- and p-Si, in agreement with the measurements. This is to be contrasted with Schottky barrier formation on III–V semiconductors, such as GaAs, where previous theoretical studies indicate that pinning is often due to surface antisite defect levels [6] which lead to different Fermi-level pinning positions for n- and p-type semiconductors.

O.F.S. wishes to thank the Arizona State Faculty Grants in Aid program for their support. We are grateful to the US Army Research Office (ARO-DAAG29-83-K-0122) and to the US Office of Naval Research (N00014-82-K-0447 and N00014-77-C-0537) for supporting this research.

References

[1] See, for example, P.S. Ho and K.N. Tu, Eds., Thin Films and Interfaces (North-Holland, New York, 1982). Reprinted from Thin Solid Films 93 (1983).
[2] J. Bardeen, Phys. Rev. 71 (1947) 717.
[3] J.M. Andrews and J.C. Phillips, Phys. Rev. Letters 35 (1975) 56.
[4] J.L. Freeouf, Solid State Commun. 30 (1980) 1059.
[5] W.E. Spicer, I. Lindau, P.R. Skeath, C.Y. Su and P.W. Chye, Phys. Rev. Letters 44 (1980) 420;
W. Mönch and H. Gant, Phys. Rev. Letters 48 (1982) 512;
H.H. Wieder, Appl. Phys. Letters 38 (1980) 170;
R.H. Williams, V. Montgomery and R.R. Varma, J. Phys. C11 (1978) L735;
R.W. Grant, J.R. Waldrop, S.P. Kowadzyk and E.A. Kraut, J. Vacuum Sci. Technol. 19 (1981) 477
[6] R.E. Allen and J.D. Dow, Phys. Rev. B25 (1982) 1423;
J.D. Dow and R.E. Allen, J. Vacuum Sci. Technol. 20 (1982) 659.
[7] C.A. Mead and W.G. Spitzer, Phys. Rev. 134 (1964) A713.
[8] O.F. Sankey, R.E. Allen and J.D. Dow, Solid State Commun. 49 (1984) 1, and to be published.
[9] G. Ottaviani, K.N. Tu and J.W. Meyer, Phys. Rev. B24 (1981) 3354.
[10] P.S. Ho and G.W. Rubloff, Thin Solid Films 89 (1982) 433.
[11] D. Cherns, G.R. Anstis, J.L. Hutchinson and J.C.H. Spence, Phil. Mag. A46 (1982) 849.
[12] G.A. Baraff and M. Schlüter, Phys. Rev. Letters 41 (1978) 892.
[13] J. Bernholc, N.O. Lipari and S.T. Pantelides, Phys. Rev. Letters 41 (1978) 895.
[14] U. Lindefelt and A. Zunger, Phys. Rev. B26 (1982) 846.
[15] D.A. Papaconstantopoulos and E.N. Economou, Phys. Rev. B22 (1980) 2903.
[16] H.P. Hjalmarson, P. Vogl, D.J. Wolford and J.D. Dow, Phys. Rev. Letters 44 (1980) 810.
[17] O.F. Sankey, to be published.
[18] D.H. Lee and J.D. Joannopoulos, Phys. Rev. B23 (1981) 4997.
[19] Y.J. Chabal, D.R. Hamann, J.E. Rowe and M. Schlüter, Phys. Rev. B25 (1982) 7598.
[20] P. Vogl, H.P. Hjalmarson and J.D. Dow, J. Phys. Chem. Solids 44 (1983) 365.
[21] E.H. Poindexter, P.J. Caplan, B.E. Deal and R.R. Razouk, J. Appl. Phys. 52 (1981) 879.
[22] M.C. Chen and D.V. Lang, Phys. Rev. Letters 51 (1983) 427.

Theory of Schottky barrier formation for transition metals on Si, Ge, diamond, and $Si_x Ge_{1-x}$ alloys

Otto F. Sankey

Department of Physics, Arizona State University, Tempe, Arizona 85287

Roland E. Allen

Department of Physics, Texas A&M University, College Station, Texas 77843

John D. Dow

Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556

(Received 31 January 1984; accepted 1 May 1984)

The heights of the Schottky barriers for various transition metals on Si, Ge, diamond, and $Si_x Ge_{1-x}$ alloys are calculated using a defect model, in which the Fermi energy is pinned by deep levels associated with interfacial dangling bonds.

PACS numbers: 73.30. + y, 73.20.Hb, 73.40.Ns

I. INTRODUCTION

When a transition metal is deposited on a Si surface, it reacts with the Si, producing a thin metallic film of transition-metal silicide.[1,2] At the same time a Schottky barrier is formed, which makes the metal/semiconductor contact non-Ohmic.[3] A remarkable fact is that the barrier heights for various different transition metals (forming silicides with a wide variety of crystal structures and stoichiometries) are all equal to within $\simeq$0.4 eV,[4,5] although the transition metals themselves have *d* levels and other features of their electronic structures that vary by electron volts. For example, Ni, Pd, and Pt, when deposited on *n*-type Si produce Schottky barrier heights differing by $\simeq$0.2 eV ($\simeq$0.63, 0.73, and 0.85 eV are the barrier heights of Ni, Pd, and Pt, respectively) despite the fact that the *s*-, *p*-, and *d*-electron energies of Ni and Pt differ by $\simeq$2, 3, and 3 eV.[6] This suggests that *the Schottky barrier height is primarily a property of the Si*, and is only weakly perturbed by the transition metal or the transition-metal silicide.

We have recently proposed that the principal experimental facts concerning Si/transition-metal-silicide Schottky barriers can be simply understood in terms of Fermi-level pinning by interfacial Si dangling bonds.[7] Simply stated, the Si dangling bond at the Si/silicide interface produces a deep level in the fundamental band gap of Si. This one-electron level, for a neutral interface, is occupied by one electron and one hole. It therefore determines or "pins" the Fermi energy at the surface, being able to accept an additional electron or hole. According to the Bardeen model of Fermi-level pinning by any surface state,[8] the surface Fermi level of the semiconductor, the metal's Fermi level, and the bulk semiconductor's Fermi level all align in electronic equilibrium. This is accomplished by electronic diffusion, which produces band bending in the semiconductor. When the alignment is accomplished (Fig. 1), there is a Schottky barrier, which for *n*-type Si has a height approximately equal to the difference between the surface conduction band edge and the surface dangling-bond deep level. For *p*-Si, the barrier height is the energy difference between the dangling-bond level and the valence band maximum. Thus, with this defect and the Fermi-level pinning model, the Schottky barrier heights for *n*-Si and *p*-Si should (approximately) add up to the band gap.

This particular explanation of the barrier heights at Si/transition-metal-silicide interfaces is especially appealing because it involves Fermi-level pinning by a native defect, and therefore makes contact with the current understanding of Schottky barrier heights in III–V semiconductors. Several authors,[9–13] Spicer in particular,[9] have espoused the notion that Fermi-level pinning by native defects determines III–V barrier heights—and a unified picture of the pinning by antisite defects in many cases and by vacancies (or other defects) in other cases is now emerging.[14,15]

In this paper we extend the idea of Fermi-level pinning by dangling bonds at Si/transition-metal–silicide interfaces to interfaces of transition metal compounds with Ge, diamond,

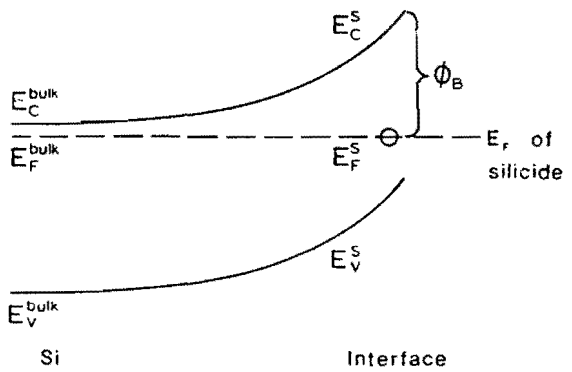


FIG. 1. Schematic illustration of Fermi-level pinning at a Si/silicide interface. Band edges for bulk Si and the Si surface, and the Fermi energies of the metal, the Si surface, and bulk Si, are all shown as functions of position. The lowest energy surface defect level that is not fully occupied (before charge is allowed to flow) is denoted by an open circle. This level approximately aligns with the Fermi levels of the *n*-type Si and the metal.

and Si_x Ge_{1-x} alloys. The model has the same features that led to its success for Si/silicide interfaces[7]: (i) The pinning defect is a dangling bond whose character is determined primarily by the semiconductor; this leads to Schottky barrier heights that are only weakly dependent on the transition metal. (ii) The formation of Schottky barriers with low coverages of transition metals is explained by the fact that the Fermi-level pinning is due to a local native defect—no appeal to metallic bulk or interface states is necessary. (iii) The observed chemical trends[16] in the barrier heights $\phi_{B,n}$ for n-type semiconductors interfaced with different transition metals are explained in terms of the weak perturbation of the dangling-bond level. (iv) The n- and p-type barrier heights add up to nearly the band gap of the semiconductor: $\phi_{B,n} + \phi_{B,p} \simeq E_{\text{gap}}$.

II. THE MODEL

We illustrate these ideas by considering a specific model of a semiconductor/transition-metal–compound interface. There are many different geometries observed for such interfaces; but here we consider a specific geometry, calculate the dangling-bond deep level, and argue that the level position is insensitive to changes in the interface geometry. For definiteness we consider Ni as the transition metal and Ge as the semiconductor, with a Ge/Ni–germanide interface. We are unaware of any detailed experimental work concerning the atomic geometry at such an interface, and so have used an interface structure (Fig. 2) identical to that reported for Si/$NiSi_2$(111) by Cherns *et al.*[17] We have a Ge "bridge atom" in the nickel–germanide connecting the semiconductor to this metallic germanide. It is bonded to three Ni atoms in the germanide, and one Ge atom at the surface of the semiconductor. The pinning defect is the dangling Ge bond that results when the Ge bridge atom is removed.

It is not necessary to fully include the nickel–germanide metallic side of the interface: Since the Schottky barrier ϕ_B in Si, and presumably also in Ge, occurs at low coverages, only the local structure of the defect is essential. Hence we can consider, instead of the dangling bond at the Ge/germanide interface, a localized defect in *bulk* Ge of a vacancy surrounded by one Ge and three Ni atoms. This is, out to second neighbors, identical to the dangling-bond defect at the true interface.

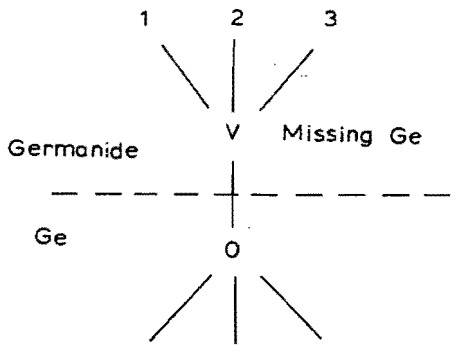


FIG. 2. An example of a Ge dangling bond at a germanide interface. Atoms 1, 2, and 3 are transition-metal atoms, atom 0 is Ge, and V is an interfacial Ge vacancy. When the vacancy is replaced by Ge, the geometry for the Ge/Ni–germanide interface is the same as that reported for $Si/NiSi_2$ (111) (Ref. 17).

Now we imagine a "cycle" in which we (i) begin with a vacancy in bulk Ge, (ii) convert three of the vacancy's nearest Ge neighbors into Ni, and (iii) finally alter the more distant neighbors on the germanide side to form semi-infinite germanide. As we have argued above, this last step is not essential, because it has very little effect on the dangling-bond deep level, and so we do not take this last step in the present work.

A. The isolated vacancy in Ge

We begin by considering the isolated vacancy in bulk Ge, and describing its deep levels using one sp^3 hybrid orbital (directed toward the vacancy) on each of its neighbors. Labeling the four neighbors 0, 1, 2, and 3, and denoting the sp^3 hybrid[18] centered on the ith neighbor and directed toward the vacancy by $|i\rangle$, we construct the A_1 symmetric (s-like) and T_2 symmetric (p-like) states of the vacancy:

$$|A_1\rangle = (1/2)(|0\rangle + |1\rangle + |2\rangle + |3\rangle),$$
$$|T_2,//\rangle = (12)^{-1/2}(3|0\rangle - |1\rangle - |2\rangle - |3\rangle),$$
$$|T_2,\perp,1,\rangle = 2^{-1/2}(|1\rangle - |3\rangle),$$
and
$$|T_2,\perp,2\rangle = 6^{-1/2}(|1\rangle + |3\rangle - 2|2\rangle).$$

Denoting the Hamiltonian matrix elements between the four hybrid orbitals surrounding the vacancy in the perfect Ge host, $\epsilon = \langle i|H|i\rangle$ for any i and $t = -\langle i|H|j\rangle$ with $j \neq i$, we find that the vacancy energies are

$$E(A_1) = \epsilon - 3t,$$
$$E(T_2) = \epsilon + t.$$

Thus, the two parameters ϵ and t can be determined uniquely from the previously calculated A_1 and T_2 *bulk* vacancy levels. Virtually all calculations[19–24] agree that the bulk vacancy level $E(A_1)$ is deep in the valence band and $E(T_2)$ lies in the fundamental band gap for all homopolar group IV semiconductors.

B. Effects of converting three Ge atoms to Ni

Now we convert three of the Ge atoms (numbers 1, 2, and 3) surrounding the vacancy into Ni atoms, in two steps: (i) we increase the sp^3 hybrid energy ϵ to $\epsilon + V$ on sites 1, 2, and 3, where we have $V \simeq 5$ eV,[25] and (ii) we introduce d levels, which are energetically deep in the valence band, only weakly coupled to the dangling bond at site 0 (either directly, through a second-neighbor interaction, or indirectly, through a ring of five nearest-neighbor interactions), and almost irrelevant (as we shall see) to the determination of the dangling-bond energy.

When V becomes nonzero, the symmetry of the defect is reduced from T_d to C_{3v}. The $|A_1\rangle$ and $|T_2,//\rangle$ states of the isolated vacancy are a_1-symmetric (σ-like) states of C_{3v}, whereas the two $|T_2,\perp\rangle$ states are e-symmetric (π-like) states. As V increases from zero to $\simeq 5$ eV, the π-bonded e-symmetric states are pushed through the gap into the conduction band—and become essentially irrelevant to the Schottky barrier formation. The two a_1 σ-bonded states of the isolated

vacancy, $|A_1\rangle$ and $|T_2,//\rangle$ are mixed by V, and are the interesting ones for the Schottky barrier problem.

Thus, the essential physics is reduced to a two-level problem, involving the effects of V on the a_1-symmetric $|A_1\rangle$ and $|T_2,//\rangle$ Ge vacancy levels. The atoms 1, 2, and 3 for $V=0$ are Ge, and their interaction is t ($\simeq 0.4$ eV). But for $V \simeq 5$ eV, the atoms 1, 2, and 3 are Ni, with hybrid energies out of resonance with the Ge, and their effective interaction with the Ge hybrid (in perturbation theory) is weak, $t^2/V \simeq 0.03$ eV.

A schematic energy level diagram is shown in Fig. 3, for various values of V. As the 1, 2, and 3 atoms are converted from semiconductor atoms into Ni (i.e., as V increases from 0 to $\simeq 5$ eV, a value between $5t$ and $20t$, see Table I), the T_2 vacancy level moves into the conduction band (and becomes irrelevant). But the A_1 vacancy level, which lies deep in the valence band, and hence has often been ignored, is driven up in energy by the electropositivity of the transition metal, until it must bend over below the bulk T_2 level (by an approximate level-crossing theorem), saturating for $V \to \infty$ at the hybrid energy ϵ. Hence, further increasing V from $\simeq 5$eV for Ni by an amount of order electron volts does not greatly alter the energy of the dangling-bond level. The dangling-bond level is "deep-level pinned" in the sense of Hjalmarson *et al.*[19]; it is semiconductorlike; and it is a deep level in the fundamental band gap capable of Fermi-level pinning.

If now we "turn on" the d-states, they press the dangling-bond level upward only slightly in energy. Estimates of this effect using Harrison's scaling rules[25] show that it is small and that the vacancy quite effectively shelters the Si dangling bond from the d-orbitals of the transition metal, even when the d orbitals are nearly resonant with the dangling bond. The d levels do play a minor role in determining the occupancy of the neutral dangling bond[7]: It has one electron and one hole—and hence, is the Fermi-level pinning defect of both n-type and p-type semiconductors.

III. RESULTS

We have used this simple model to predict the surface dangling-bond deep levels of Si, Ge, diamond, and Si_x Ge_{1-x} interfaced with their transition metal compounds. (We have also confirmed that the essential physics of the model is correct by performing extensive and rigorous calculations for a cluster of a vacancy and three Ni atoms in Si, and by explicitly treating an interface between semi-infinite slabs of $NiSi_2$ and Si.[26]) We first determine ϵ and t by fitting calculated bulk vacancy levels. Our results are presented in Table I. The matrix elements V are extracted from tables of atomic orbital energies.[25] The deep levels associated with the dangling bonds at various interfaces are then computed in terms of ϵ, t, and V.

Slight variations in V for different transition metals produce weak variations in ϕ_B, as expected [see Fig. 3(b)]. The chemical trends with respect to such variations[1,2,16] are explained by the theory.

Although the predictions of the theory for Schottky barrier heights are gratifyingly close to the measured barrier heights, the theory in its original form is limited by the ~ 0.3 eV uncertainty in even the best theoretical predictions of the bulk vacancy levels. To circumvent this problem we adjust ϵ to fit the interfacial Fermi level and Schottky barrier heights for the $Si/NiSi_2(111)$ interface, and then use the predicted changes in ϵ from Si to Ge to predict the variation in the interfacial Si_x Ge_{1-x}/nickel–silicide–germanide dangling-bond level with alloy composition x. The results are given in Fig. 4, and rather dramatically account for the observed chemical trend in the Schottky barrier height,[27] assuming pinning by the dangling-bond deep level.

TABLE I. The calculated A_1- and T_2-symmetric energy levels for the unrelaxed Si, Ge, and C vacancies from tight-binding theory. The dangling-bond energy ϵ and dangling-bond interaction t are extracted from the A_1- and T_2-symmetric levels. All energies are in eV, and the top of the valence band is defined to be the zero of energy. Other values of ϵ for Si, determined from reported calculations of the Si vacancy, range from 0.09 to 0.45 eV.

	$E(A_1)$	$E(T_2)$	t	ϵ
Si[a]	−1.10	0.48	0.40	0.09
Ge[a]	−0.98	0.04	0.26	−0.22
C[b]	−0.65	2.43	0.77	1.66

[a] Reference 24.
[b] Reference 19.

The theory also predicts a dangling-bond deep level and a Schottky barrier height as a result of transition-metal depositions on diamond. We are not aware of any data for this system, but we note that diamond/Al, /Au, and /Ba interfaces have been studied experimentally,[28,29] and should also be expected to have dangling-bond states. The Schottky barrier height obtained, assuming Fermi-level pinning by diamond dangling bonds, is in excellent agreement with the data (Fig. 5).

IV. SUMMARY

It should be emphasized that we have presented a simple model which is meant to display the essential physics of Fermi-level pinning by interfacial Si dangling bonds. The defect

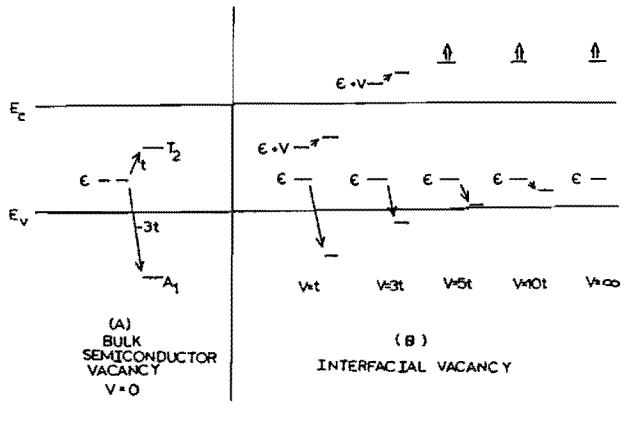


FIG. 3. (a) Schematic energy level diagram of the A_1- and T_2-symmetric levels of a vacancy in an elemental semiconductor. E_c and E_v are the conduction and valence band edges. (b) The a_1-symmetric levels for an interfacial vacancy surrounded by one semiconductor atom with orbital energy ϵ and 3 transition-metal atoms with hybrid energy $\epsilon + V$. The levels for values of V equal to t, $3t$, $5t$, $10t$, and ∞ are shown. Note that for $V >> t$, the band-gap level becomes "pinned" to the semiconductor dangling-bond energy ϵ. Double arrows indicate that the $\epsilon + V$ level in the conduction band is off the figure.

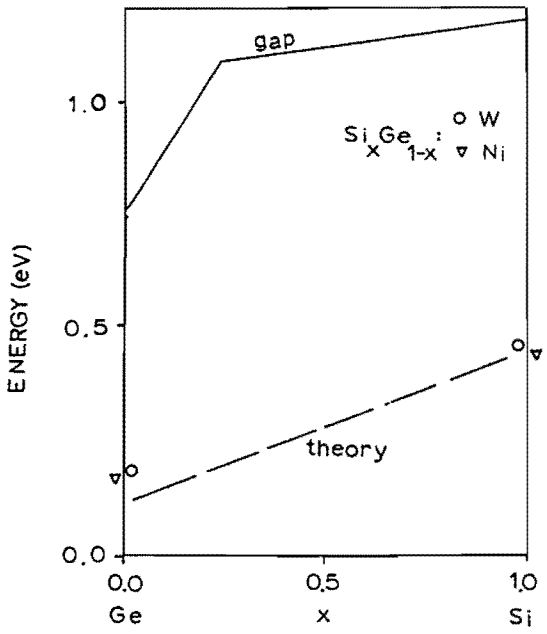


FIG. 4. Predicted Fermi level in the Fermi-level pinning model due to dangling-bond defects in Si_x Ge_{1-x} alloys. Because of the theoretical uncertainties in determining the dangling-bond energy, the theory has been fit to the data for Si. The data for Si are from, e.g., Ref. 4, and the data for Ge are from Ref. 27.

levels that we predict may be inaccurate by a few tenths of eV, and are undoubtedly inhomogeneously broadened on a $\simeq$0.1 eV scale, due to varying local environments.[30]

The model is based on the idea of Fermi-level pinning by defect levels, and therefore provides a unified explanation of Schottky barrier heights for Group-IV as well as III–V semiconductors. No theory of Schottky barrier formation is universally accepted yet, but the defect model is by far the most

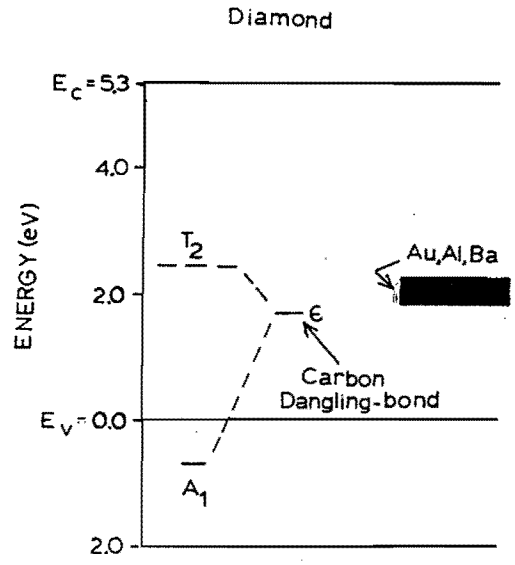


FIG. 5. The A_1 and T_2 levels of a C vacancy in diamond, after Ref. 19, and the dangling-bond energy ϵ for an interfacial vacancy. The data for (nontransition-metal) Schottky barrier heights on p-type diamond are shown for comparison (Ref. 28 and 29).

widely accepted, and the only general theory for the technologically important semiconductors that has not been manifestly disproven by numerous experiments. This is not to say that defect mechanisms are the only means of Schottky barrier formation;[31,32] in some cases metal/semiconductor interfaces might be produced without very many defects, in which case metal-induced gap states[33] might play a role in determining the Schottky barriers. However, for years metal-induced gap-state theories have been widely viewed as incapable of providing a satisfactory general description of the common III–V barrier heights, because they can consistently explain at most very few experiments. For example, data demonstrating that (i) the n-GaAs Schottky barrier can be annealed away at the antisite defect annealing temperature while the p-GaAs barrier remains,[34] and that (ii) the n-InP barrier height switches from being $\simeq$0.1 eV for reactive metals to $\simeq$0.5 eV for nonreactive metals[35] are easily explained by a defect model but not by a metal-induced gap-state model.

The simple theory presented here accounts well for the major observations concerning Schottky barriers resulting from transition-metal deposition on group-IV semiconductors: (i) Schottky barriers form at submonolayer coverages because the Fermi-level pinning defect is localized; (ii) Schottky barrier heights exhibit only weak dependences on the transition metals, because the dangling-bond level is deep-level pinned; (iii) details of the crystal structure, stoichiometry, and interface geometry are minor perturbations on the surface dangling-bond deep level, because the primary role of the silicide or germanide is merely to supply a repulsive potential that pushes the deep dangling-bond level back into the semiconductor; (iv) the Schottky barrier heights for n-type and p-type group–IV semiconductors add up to $\simeq E_{gap}$ because the dangling bond level is both a donor and an acceptor; (v) the variations of Schottky barrier heights depend on the chemistry of the semiconductor, as displayed in the dependence on x of the Si_x Ge_{1-x} barrier height.

ACKNOWLEDGMENTS

We gratefully acknowledge the generous support of the U. S. Army Research Office (Contract No. DAAG29-83-K-0122). O. F. Sankey acknowledges an Arizona State Faculty Grant in Aid.

[1]See, for example, *Thin Films and Interfaces*, edited by P. S. Ho and K. N. Tu (North-Holland, New York, 1982). Reprinted from Thin Solid Films 93 (1983).
[2]P. S. Ho and G. W. Rubloff, Thin Solid Films 89, 433 (1982).
[3]E. H. Rhoderick, *Metal–Semiconductor Contacts* (Clarendon, Oxford, 1978).
[4]J. M. Andrews and J. C. Phillips, Phys. Rev. Lett. 35, 56 (1975).
[5]J. L. Freeouf, Solid State Commun. 30, 1059 (1980).
[6]O. Bisi and C. Calandra, J. Phys. C 14, 5479 (1981).
[7]O. F. Sankey, R. E. Allen, and J. D. Dow, Solid State Commun. 49, 1 (1984).

[8]J. Bardeen, Phys. Rev. **71**, 717 (1947).
[9]W. E. Spicer, I. Lindau, P. R. Skeath, C. Y. Su, and P. W. Chye, Phys. Rev. Lett. **44**, 420 (1980).
[10]W. Mönch and H. Gant, Phys. Rev. Lett. **48**, 512 (1982).
[11]H. H. Wieder, Appl. Phys. Lett. **38**, 170 (1980).
[12]R. H. Williams, Surf. Sci. **32**, 122 (1983).
[13]M. S. Daw and D. L. Smith, Phys. Rev. B **20**, 5150 (1979).
[14]R. E. Allen and J. D. Dow, Phys. Rev. B **25**, 1423 (1982); J. D. Dow and R. E. Allen, J. Vac. Sci. Technol. **20**, 659 (1982).
[15]R. E. Allen, T. J. Humphreys, J. D. Dow, and O. F. Sankey, J. Vac. Sci. Technol. B (these proceedings).
[16]G. Ottaviani, K. N. Tu, and J. W. Meyer, Phys. Rev. B **24**, 3354 (1981).
[17]D. Cherns, G. R. Anstis, J. L. Hutchison, and J. C. H. Spence, Philos. Mag. A **46**, 849 (1982).
[18]S. Y. Ren, W. M. Hu, O. F. Sankey, and J. D. Dow, Phys. Rev. B **26**, 951 (1982).
[19]H. P. Hjalmarson, P. Vogl, D. J. Wolford, and J. D. Dow, Phys. Rev. Lett. **44**, 810 (1980); P. Vogl, H. P. Hjalmarson and J. D. Dow, J. Phys. Chem. Solids **44**, 365 (1983).
[20]G. A. Baraff and M. Schlüter, Phys. Rev. Lett. **44**, 892 (1978).
[21]J. Bernold, N. O. Lipari, and S. T. Pantelides, Phys. Rev. Lett. **41**, 895 (1978).
[22]U. Lindefelt and A. Zunger, Phys. Rev. B **26**, 846 (1982).
[23]D. A. Papaconstantopoulous and E. N. Economou, Phys. Rev. B **22**, 2903 (1980).
[24]K. F. Newman and J. D. Dow, *Theory of Deep Impurities in Silicon/Germanium Alloys* (to be published).
[25]W. A. Harrison, *Electronic Structure and the Properties of Solids* (Freeman, San Francisco, 1980).
[26]O. F. Sankey (to be published).
[27]E. Y. Chan and H. C. Card, IEEE Trans. Electron Devices ED-27, 78 (1980).
[28]C. A. Mead, Solid State Electron. **9**, 1023 (1966).
[29]C. A. Mead and T. C. McGill, Phys. Lett. A **58**, 249 (1976).
[30]This is very likely the primary source of broadening. Extended metallic conduction band states are only weakly coupled to the Si dangling band either through a second-neighbor interaction or indirectly through a ring of five nearest-neighbor interactions.
[31]J. L. Freeouf and J. M. Woodall, Appl. Phys. Lett. **39**, 727 (1981); J. L. Freeouf, Solid State Commun. **30**, 1059 (1980); J. M. Woodall, G. D. Petit, T. N. Jackson, C. Lanza, K. L. Kavanagh, and J. W. Mayer, Phys. Rev. Lett. **51**, 1783 (1983).
[32]R. Ludeke and L. Esaki, Phys. Rev. Lett. **33**, 653 (1974); R. Ludeke and A. Koma, *ibid.* **34**, 817 (1975); **39**, 1042 (1977); and Surf. Sci. **132**,143 (1983).
[33]V. Heine, Phys. Rev. A **138**, 1689 (1965); J. C. Inkson, J. Phys. C **6**, 1350 (1973); E. Louis, F. Yndurain, and F. Flores, Phys. Rev. B **13**, 4408 (1976); S. G. Louie, J. R. Chelikowsky, and M. L. Cohen, *ibid.* **15**, 2154 (1977); C. Tejedor, F. Flores, and E. Louis, J. Phys. C **10**, 2163 (1977); F. Guinea, J. Sanchez-Dehesa, and F. Flores, *ibid.* **16**, 6499 (1983); J. Tersoff, Phys. Rev. Lett. **52**, 465 (1984); J. Sanchez-Dehesa and F. Flores, Solid State Commun. **50**, 29 (1984).
[34]W. Mönch, Surf. Sci. **132**, 92 (1983); (to be published).
[35]R. H. Williams, V. Montgomery, and R. R. Varma, J. Phys. C **11**, L735 (1978); R. H. Williams and M. H. Patterson, Appl. Phys. Lett. **40**, 484 (1982); R. H. Williams, Surf. Sci. **132**, 122 (1983); J. D. Dow and R. E. Allen, J. Vac. Sci. Technol. **20**, 659 (1982).

PHYSICAL REVIEW B VOLUME 32, NUMBER 2 15 JULY 1985

Theoretical investigation of the pressure dependences of energy gaps in semiconductors

Seongbok Lee, J. Sanchez-Dehesa,* and John D. Dow
Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556
(Received 10 December 1984)

The observed dependences on pressure of the energy gaps of Si, Ge, and GaAs at symmetry points in the Brillouin zone are successfully calculated using a variational method based on density-functional theory. The negative pressure derivatives of the gaps at the X point of the conduction band relative to the valence-band maxima are due to the d states.

I. INTRODUCTION

In this paper we report successful *a priori* calculations of the pressure dependences of the band gaps of Si, Ge, and GaAs. The only inputs to our calculations are the crystal structures of the materials and the local pseudopotentials of the atomic constituents. The theory is based on a variational minimization of the total energy using a Wannier-function basis and a density-functional formalism applied self-consistently to the valence electrons.[1,2] Our calculated pressure dependences of the band gaps are in good agreement with data; this suggests that local-density theory accurately predicts the changes with pressure of the band gaps despite the fact that in its present form it yields poor predictions for the zero-pressure band gaps themselves.

Previous theoretical studies of pressure-dependent band gaps have employed the self-consistent orthogonalized-plane-wave (OPW) method,[3] various semiempirical schemes (such as those that fit the pressure dependences of pseudopotential form factors[4,5] or empirical tight-binding parameters[4] to data), or a modified version of Van Vechten's dielectric theory.[6] To our knowledge, no successful *a priori* theory has been reported previously.

II. METHOD

Following the fundamental work of Tejedor and Verges,[1,2] we minimize the total energy of the valence electron system, with respect to the parameters β of basis Slater orbitals. The relevant s and p orbitals for the electrons of the nth atomic shell have the following radial dependences:

$$F_s(r,\beta_s)=c_s r^{n-1}\exp(-\beta_s r)$$

and

$$F_p(r,\beta_p)=c_p r^{n-1}\exp(-\beta_p r)\ ,$$

where we have

$$c_s=(2\beta_s)^{(2n+1)/2}/\sqrt{(2n)!}$$

and

$$c^p=(2\beta_p)^{(2n+1/2)}/\sqrt{(2n)!}\ .$$

From these orbitals we construct the sp^3 hybrid orbitals[8]

$$\phi^a_{vm}(\mathbf{r})=[1/(4\sqrt{\pi})][F_s(r,\beta^a_s)+\sqrt{3}(\mathbf{v}_m\cdot\mathbf{r}/r)F_p(r,\beta^a_p)]\ ,$$

$$\phi^c_{-vm}(\mathbf{r})=[1/(4\sqrt{\pi})][F_s(r,\beta^c_s)+\sqrt{3}(\mathbf{v}_m\cdot\mathbf{r}/r)F_p(r,\beta^c_p)]\ ,$$

and the bonding combination of the hybrids[9]

$$f^v_m(\mathbf{r},\beta)=(\sin\theta)\phi^a_{vm}(\mathbf{r})$$
$$=(\cos\theta)\phi^c_{-vm}(\mathbf{r}-\mathbf{v}_m a/4)\ .$$

Using the bonding combination[10] and employing Löwdin's symmetrical orthogonalization method we construct $\Psi_{m\mathbf{k}}(\mathbf{r},\beta)$, a set of Bloch-type linear combinations of the orthonormalized functions, from which we can obtain the Wannier functions,

$$w_m(\mathbf{r})=N^{-1/2}\sum_{\substack{\mathbf{k}\\ \mathrm{BZ}}}\Psi_{m\mathbf{k}}(\mathbf{r},\beta)$$

[where the sum is over the Brillouin zone (BZ)] and the local charge density

$$\rho(\mathbf{r},\beta)=2\sum_m\sum_{\mathbf{r}} w^*_m(\mathbf{r}-\mathbf{R})w_m(\mathbf{r}-\mathbf{R})\ .$$

The charge density is a function of the Slater parameters β and the position $\mathbf{r}$. We evaluate the total energy, which consists of the ion-ion interaction energy E_1, the kinetic energy T, the electron-ion interaction E_2, the Hartree energy E_3, and the exchange-correlation energy E_{xc}, following established procedures.[11]

The total energy is varied (numerically) with respect to the parameters β until a minimum is found. The critical values of β determine the ground-state charge density and can be used to construct the local-density Hamiltonian.[12] Diagonalization of this Hamiltonian using 89 plane waves as a basis produces an approximation to the energy band structure $\epsilon_m(\mathbf{k})$, for $m=1,2,\ldots,89$.

To determine the hydrostatic pressure dependences of the band structures this method is repeated for several different lattice constants.

III. INDEPENDENCE OF THE PSEUDOPOTENTIAL

This approach is meaningful only if it gives results which do not depend sensitively on the choice of pseudopotential from among those considered to be "good." To verify that this is indeed the case we have executed the calculations for Si using three different pseudopotentials: (i) the Hamann-Schlüter-Chiang first-principles pseudopo-

TABLE I. Calculated pressure derivatives dE_{gap}/dp (in meV/kbar) of the band gaps of Si using different local pseudopotentials.

Gap[a]	First principles[b]	Hard core[c]	Soft core[d]	Expt.	Semiempirical calculation
$\Gamma_c-\Gamma_v$	0.48	0.48	0.34	1 ± 1[e]	1.3,[f] 3.7[g]
L_c-L_v	4.45	4.35	4.41	$6.2+0.4$[e]	6.6,[f] 2.7[g]
X_c-X_v	1.64	1.59	1.26	3.0	3.6,[f] 3.6[g]
$L_c-\Gamma_v$	3.30	3.21	3.23		5.5,[f] 1.2[g]
$X_c-\Gamma_v$	-1.34	-1.42	-1.86	-1.5	0.5,[f] -0.1[g]

[a] $X_c-\Gamma_v$ means the gap from the top of the valence band at Γ to the conduction band at X.
[b] Reference 13.
[c] Reference 14.
[d] Reference 15.
[e] Reference 20.
[f] Reference 5.
[g] Reference 7.

TABLE II. Pressure derivatives dE/dp (in meV/kbar) of the band gaps for GaAs and Ge using the "soft-core" ionic pseudopotential of Ref. 27.

	GaAs			Ge		
Gap	Our result	Expt.	Semiempirical calculations	Our result	Expt.	Semiempirical calculations
$\Gamma_c-\Gamma_v$	10.50	10.74,[a] 12.6[b] 10.7–11.7[c]	11,[d] 13.3[e]	16.19	15.3,[f] 11 14.2[g]	14.3,[d] 16.2[e]
L_c-L_v	4.46	5.0	4.5,[d] 7.4[e]	6.28	7.5	7.1,[d] 8.8[e]
X_c-X_v	0.78		3.6,[d] 4.6[e]	2.36	5.5	4.4,[d] 5.4[e]
$L_c-\Gamma_v$	2.93		2.8,[d] 6.2[e]	4.90	5.0	8.0,[f] 5.4,[d] 6.6[e]
$X_c-\Gamma_v$	-2.52	-1.34[a]	-0.8,[d] 1.5[e]	-1.11	-1.5	-0.1,[d] 2.7[e]

[a] D. J. Wolford and J. E. Bradley, Solid State Commmun. (to be published).
[b] Reference 23.
[c] As compiled in Ref. 22.
[d] Reference 7.
[e] Reference 5.
[f] Reference 22.
[g] Reference 24.

TABLE III. Nonlinear pressure coefficients $\frac{1}{2}d^2E_0/dp^2$ (in eV/kbar^2) of the direct gap of GaAs and Ge.

	Theory	Expt.
Ge	-7.04×10^{-5}	$-(4.5\pm1)\times10^{-5}$
GaAs	-1.88×10^{-5}	-3.77×10^{-5}

TABLE IV. Pressure derivatives dE_{gap}/dp (in meV/kbar) of the band gaps of Ge using different basis sets of localized orbitals in order to diagonalize the Hamiltonian.

Gap	$4s$ $4p$ $5s$ (10 bands)	$4s$ $4p$ $4d$ (18 bands)	$4s$ $4p$ $4d$ $5s$ (20 bands)
$\Gamma_c-\Gamma_v$	16.01	17.89	16.14
L_c-L_v	11.63	5.74	5.95
X_c-X_v	7.50	2.33	2.07
$L_c-\Gamma_v$	9.83	4.67	4.64
$X_c-\Gamma_v$	3.62	-0.53	0.36

tential,[13] (ii) the Harris-Jones "hard-core" pseudopotential,[14] and (iii) the Schlüter *et al.* "soft-core" empirical pseudopotential,[15] which was adjusted to fit the Si band structure at zero pressure. Our results are given in Table I where the connection between pressure and lattice constant is supplied by Murnaghan's equation of state[16]

$$p=[B_0/dB_0/dp)][(a_0/a)^{3dB_0/dp}-1] .$$

Here B_0 is the bulk modulus of Si (978.8 kbar) and dB_0/dp is its pressure derivative (4.24).[17]

The agreement among the pressure dependences of the gaps as computed using the three pseudopotentials is excellent,[18] even for the gaps from any symmetry point of the valence band to the conduction band at X, where previous calculations with the empirical pseudopotential have generally failed to produce even the correct sign for $dE_{\rm gap}/dp$. This success causes us to extend Hamann's[19] conjecture—that different "good" pseudopotentials give roughly the same band gaps (in local-density theory)—to include pressure derivatives as well. However, unlike the absolute band gaps, which are too small in local-density theory, the pressure derivatives are predicted rather well.

Our predictions for the pressure-dependences of the gaps of Ge and GaAs (using the pseudopotentials of Ref. 27) are given in Table II, and are in good agreement with the data.[21]

We have also computed the second derivatives of the band gaps of Ge and GaAs and find that in these materials the second derivatives of the direct band gaps are sizable (see Table III) as observed by Welber *et al.*[22,23] for GaAs and Ge.[25]

The reason that $dE_{\rm gap}/dp$ is negative for the gap between the valence band at Γ to the conduction band at X is the strong influence of the d levels that lie in energy well above the X minima of the conduction band. These levels repel the conduction band at X, forcing it downward in energy (relative to the minimum at Γ); without the d states the pressure dependence of the X conduction-band minima is not correctly reproduced by the theory. This is demonstrated in Table IV for Ge, where we display $dE_{\rm gap}/dp$ as computed with and without d orbitals, in models with 10, 18, and 20 basis orbitals per unit cell. The empirical pseudopotential method does not adequately represent the effects of the d states, and hence does not predict the correct sign for $dE_{\rm gap}/dp$ at X.

IV. CONCLUSION

Hence we conclude that the derivatives of the band gaps of Si, Ge, and GaAs, and probably other semiconductors, can be predicted accurately using the variational Tejedor-Verges localized-orbital method. The negative value of $dE_{\rm gap}/dp$ for the gap between the valence-band maximum and the conduction band at the X point of the Brillouin zone is attributable to high-energy d states that depress the X minima.

The success of the theory in computing pressure derivatives of band gaps, despite the fact that all local-density theories to date have predicted absolute band-gap energies in error by typically 50%, is reassuring. This indicates that the localized-orbital method can be used to study the pressure dependences of deep impurity levels and the band gaps of strained superlattices.

Finally, since the theory predicts pressure dependences of absolute band gaps in good agreement with the data, it also implies that the corrections[26] to local-density theory necessary to produce the observed band gaps necessarily must be volume and pressure independent—in order to preserve the agreement between local-density theory and data, as found here.

ACKNOWLEDGMENTS

We are grateful to K. E. Newman for stimulating conversations, and to D. J. Wolford for communicating his data prior to publication. This work was supported by the U. S. Army Research Office (Contract No. DAAG29-83-K-0122).

---

*Permanent address: Departamento de Fisica del Estado Solido, Universidad Autónoma de Madrid, Madrid 28034, Spain.

[1]C. Tejedor and J. A. Verges, Phys. Rev. B 19, 2283 (1979).

[2]J. A. Verges and C. Tejedor, Phys. Rev. B 20, 4251 (1980).

[3]F. Herman, R. Kortum, C. D. Kuglin, L. Goroff, and L. Kleinman, Phys. Rev. 132, 1524 (1963).

[4]P. J. Melz, J. Phys. Chem. Solids 28, 1441 (1967).

[5]Y. F. Tsay, S. S. Mitra, and B. Bendow, Phys. Rev. B 10, 1476 (1976).

[6]S. Y. Ren, J. D. Dow, and D. J. Wolford, Phys. Rev. B 25, 7661 (1980).

[7]D. L. Campausen, G. A. Neville Connell, and W. Paul, Phys. Rev. Lett. 26, 184 (1971).

[8]The tetrahedral vectors v_m are $(1,1,1)$, $(-1,1,-1)$, $(1,-1,-1)$, $(-1,-1,1)$, and the superscripts a and c denote anion and cation, respectively.

[9]If we take the anion site as the origin, the four first neighbors are at the positions $v_m a/4$ $(m=1,2,3,4)$, where a is the lattice parameter.

[10]The set of variational parameters β in f_m contains θ, β_s^a, β_p^a, β_s^c, β_p^c. It must be emphasized that in homopolar semiconductors we have $\phi^a=\phi^c$, which implies that we have $\theta=\pi/4$, so the variational parameters β reduce to only two parameters β_s and β_p.

[11]See Eqs. (3) to (7) in Ref. 2.

[12]See, for example, Eq. (2.2) in Ref. 13.

[13]The intrinsically nonlocal psuedopotential is approximated by the local $l=0$ s psuedopotential as fitted to an analytical form by G. Bachelet, D. R. Hamann, and M. Schlüter, Phys. Rev. B 26, 4199 (1982).

[14]J. Harris and R. O. Jones, Phys. Rev. Lett. 41, 191 (1978).

[15]M. Schlüter, J. R. Chelikowsky, S. G. Louie, and M. L. Cohen, Phys. Rev. B 12, 4200 (1975).

[16]F. D. Murnaghan, Proc. Natl. Acad. Sci. U.S.A. 30, 244 (1944).

[17]The parameters of the equation of state used for Si, Ge, and GaAs are taken from H. J. McSkimin, J. Appl. Phys. 24, 988 (1953); 34, 651 (1963); 35, 2161 (1964).

[18]To obtain the second derivatives, we have calculated the band structures of Si at five different lattice constants up to 5.267 Å, corresponding to a maximum pressure of 109 kbar. Then the calculated gaps were least-squares fitted to a second-order polynomial in p (pressure), which was differentiated. The nonlinear terms are negligible in this particular case.

[19]D. R. Hamann, Phys. Rev. Lett. 42, 662 (1979).

[20]E. Schmidt and K. Vedam, Solid State Commun. 9, 1187 (1971).

[21]Our smallest lattice constants for these materials are 5.49 and 5.433 Å, corresponding to maximum pressures of 85 and 123 kbar for Ge and GaAs, respectively. Note that the critical pressures of phase transformation for Ge and GaAs are 105 and 180 kbar, respectively.

[22]B. Welber, M. Cardona, Y. F. Tsay, and B. Bendow, Phys. Rev. B 15, 875 (1977).

[23]B. Welber, M. Cardona, C. K. Kim, and S. Rodriguez, Phys. Rev. B 12, 5729 (1975).

[24]P. J. Melz, J. Phys. Chem. Solids 32, 209 (1971).

[25]As stressed by these authors, part of this nonlinearity is a consequence of the nonlinearity of the change of lattice parameters with pressure.

[26]M. Schlüter and L. J. Sham, Bull. Am. Phys. Soc. 29, 351 (1984).

[27]W. E. Pickett, S. G. Louie, and M. L. Cohen, Phys. Rev. B 17, 815 (1978).

PHYSICAL REVIEW B VOLUME 31, NUMBER 6 15 MARCH 1985

Theory of charge-state splittings of deep levels

Seongbok Lee and John D. Dow
Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556

Otto F. Sankey
Department of Physics, Arizona State University, Tempe, Arizona 85287

(Received 13 July 1984)

The Green's-function method, with an empirical tight-binding basis, is used to determine the deep levels of the singly ionized and neutral impurities S, Se, and Te in Si. The impurity potentials are determined self-consistently. The resulting theory accounts for the observed charge-state splittings of neutral and singly ionized A_1 deep levels, obtaining, for S, Se, and Te, 0.23, 0.22, and 0.19 eV (to be compared with experimental values of 0.30, 0.29, and 0.21 eV, and with a self-consistent local-density-theory value for S of 0.20).

I. INTRODUCTION

The ionization energy of a "deep" impurity in a semiconductor is defined as the energy required to remove an electron (hole) from the occupied deep level in the band gap to the conduction- (valence-) band edge, and depends on the charge state of the impurity—namely, whether it is initially neutral, or charged either positively or negatively, with an integral multiple of the proton's charge. A deep impurity level is one produced by the impurity's central-cell potential.[1,2] The charge-state splitting of a deep impurity level in the band gap is the difference between the ionization energies of the impurity with charge Q and the impurity with one fewer electron (or hole). Experimentally, for defects in covalent semiconductors, charge-state splittings are typically a few tenths of an eV: For example, the ionization energy of S^+ in Si (0.613 eV) is 0.295 eV larger than the ionization energy for S^0 in Si (0.318 eV).[3,4] The purpose of this paper is to account for chemical trends in the observed charge-state splittings of S, Se, and Te in Si.

The charge-state splitting of a deep level is a many-body effect which results from the Coulomb interactions among electrons. In a potential-scattering one-electron theory, the type of theory one normally uses when plotting a one-electron energy-level scheme or band structure, the defect potential is state independent and the charge-state splitting is identically zero. In one-electron theories of the change-of-mean-field type, such as Hartree or Hartree-Fock theory, the charge-state splitting is nonzero, because the one-electron potentials are state dependent. To illustrate this point, consider atomic He in the Hartree approximation, assuming $1s$ orbital wave functions of the form $\exp(-Zr/a)$, where Z is the effective charge and is treated as a variational parameter. For the $(1s)^2$ neutral ground state, we have an effective charge $Z=\frac{27}{16}$ and $\epsilon_{1s}(Z=\frac{27}{16})=-\frac{459}{256}$Ry, and for the singly ionized state, we have $Z=2$ and $\epsilon_{1s}(Z=2)=-4$ Ry.[5] The ionization energy E_I^+ of He^+ is 4 Ry or 54.40 eV. The ionization energy E_I^0 of He^0 is

$$E_I^0=E_{\rm tot}(He^+)-E_{\rm tot}(He^0)$$
$$=\epsilon_{1s}(Z=2)-2\epsilon_{1s}(Z=\tfrac{27}{16})+U$$
$$=(\tfrac{217}{128})\ \text{Ry},$$

where

$$U=\langle 1s,1s\,|\,e^2/r\,|\,1s,1s\rangle=(\tfrac{135}{64})\ \text{Ry}$$

is the electron-electron repulsion integral. The charge-state splitting of atomic He and He^+, calculated in this Hartree approximation, is 2.30 Ry:

$$\Delta E=E_I^+-E_I^0$$
$$=-2\epsilon_{1s}(Z=2)+2\epsilon_{1s}(Z=\tfrac{27}{16})-U=(\tfrac{295}{128})\ \text{Ry}.$$

Experimentally ΔE is 2.19 Ry,[6] indicating that exchange and correlation effects contribute of order 5% (and presumably can be neglected in calculations of charge-state splittings for defects in solids).

For S, Se, and Te in vacuum, the experimental charge-state splittings between the neutral (s^2p^4 configuration) and singly ionized (s^2p^3) states are 13.04, 11.75, and 9.59 eV, respectively.[6] However, the corresponding charge-state splittings for the deep levels associated with S, Se, and Te in the fundamental gap of Si are approximately *two orders of magnitude smaller*: 0.295, 0.286, and 0.212 eV.[3,4,7] These small splittings, which are typical of deep levels in semiconductors, are in accord with theoretical predictions.[8] Roughly speaking, one of these orders of magnitude comes from the screening of the Coulomb interactions, $e^2/\epsilon r_{\rm av}$, in the semiconductor by the dielectric polarization of the valence band, $\epsilon\simeq 12$; the second order of magnitude is attributable to the fact that the average separation of correlated electrons occupying the deep level, $r_{\rm av}$, is approximately a lattice constant $\simeq 5.43$ Å rather than a Bohr radius, 0.53 Å. The electrons are separated by such a large distance because the deep-level state in the gap is antibonding and hostlike,[1,2] with its wave function spread primarily over the four neighbors to the chalcogen impurity. Thus the average separation between electrons

is roughly the distance between Si neighbors on opposite sides of the chalcogen. (Recall that, in contrast to the hostlike antibonding deep levels, the hyperdeep levels[1,2] of S, Se, and Te, which should lie $\simeq$ 15 eV below the valence-band maximum, are the bonding chalcogenlike states.)

In this paper we present an empirical tight-binding theory of charge-state splittings for deep impurity levels in semiconductors and apply it to S, Se, and Te in Si. Hence this work is complementary to earlier studies, using local-density theory, of the S defect in Si.[8]

II. CALCULATIONS

Our model employs the Hjalmarson model of deep impurity levels,[1,2] the Vogl sp^3s^* empirical tight-binding model of electronic structure,[9] and the Haldane-Anderson model of Coulomb effects.[10] Although the Haldane-Anderson model was originally designed to treat transition metals, it has been adapted by Sankey and co-workers for the treatment of deep levels associated with interstitial defects in Si.[11] Since it provides an especially simple and convenient scheme for executing self consistent calculations in a tight-binding basis, we use it to study the charge-state splittings of substitutional deep-impurity levels, including the chalcogens in Si.

The Schrödinger equation for the perfect crystal is

$$H^0\psi_{n\vec{k}}=E_{n\vec{k}}\psi_{n\vec{k}} , \tag{1}$$

where $E_{n\vec{k}}$ is an energy band structure and $\psi_{n\vec{k}}$ are Bloch functions. A point defect or impurity breaks the translational invariance of the perfect crystal and induces some perturbation potential V. The eigenvalue equation for the imperfect solid is

$$H\psi=(H^0+V)\psi=E\psi . \tag{2}$$

The formal solution of Eq. (2) in the forbidden band gap is given by

$$\psi=G^0(E)V\psi , \tag{3}$$

where $G^0(E)=(E-H^0)^{-1}$ is the Green's operator and is real in the band gap. A nontrivial solution of Eq. (3) for ψ exists if

$$\det[1-G^0(E)V]=0=\det\left[1-P\int_{\infty}^{-\infty}dE'(E-E')^{-1}\sum_{n\vec{k}}|\psi_{n\vec{k}}\rangle\delta(E'-E_{n\vec{k}})\langle\psi_{n\vec{k}}|V\right] . \tag{4}$$

Here P denotes a principal value integral. The perfect crystal Hamiltonian H^0 in the nearest-neighbor empirical tight-binding sp^3s^*-basis model of Vogl *et al.*[9] is

$$H^0=\sum_{i,\sigma,\vec{R}}[\,|ia\sigma\vec{R}\rangle E_{i,a}\langle ia\sigma\vec{R}|+|ic\sigma\vec{R}+\vec{d}\rangle E_{i,c}\langle ic\sigma\vec{R}+\vec{d}|\,] + \sum_{i,j,\sigma,\vec{R},\vec{R}'}[\,|ia\sigma\vec{R}\rangle V_{ij}(\vec{R},\vec{R}'+\vec{d})\langle jc\sigma\vec{R}'+\vec{d}|+\text{H.c.}] . \tag{5}$$

Here $i=s$, p_x, p_y, p_z, or s^* labels the orbitals, a and c denote anion and cation (for a polar semiconductor), σ is the spin ($\uparrow$ or $\downarrow$), $\vec{R}$ specifies the unit cell, and H.c. stands for Hermitian conjugate. The transfer-matrix elements V_{ij} are nonzero only between nearest neighbors. The states $|ia\sigma\vec{R}\rangle$ and $|ic\sigma\vec{R}+\vec{d}\rangle$ are localized orbitals centered on the anion at $\vec{R}$ and the cation at $\vec{R}+\vec{d}$, respectively.[9,12] The defect potential for a single impurity located at $\vec{R}=\vec{0}$ (taken here to be an "anion" site) can be written as

$$V=\sum_{i,\sigma}|ia\sigma\vec{0}\rangle V_{ia\sigma}\langle ia\sigma\vec{0}| + \sum_{i,\sigma,\vec{R}}|ic\sigma\vec{R}+\vec{d}\rangle V_{ic\sigma}\langle ic\sigma\vec{R}+\vec{d}| , \tag{6}$$

where the sum on $\vec{R}$ is taken over four neighbors, and the basis orbitals at $\vec{0}$ are impurity orbitals. The impurity potential includes a central-cell part, because that part of the potential is responsible for the formation of deep levels; it contains a first-neighbor contribution because the impurity wave function is located on the neighbors and neglect of this part of the potential would lead to charge-state splittings too small by a factor of $\simeq$10. The off-diagonal matrix elements of the impurity potential are assumed to be independent of the ionicity of the defect, in accord with Harrison's rule[13] that they depend only on the bond length. For simplicity, longer-ranged contributions to V are neglected.

The problem of determining the deep-level energy E for a given charge state has two parts: (i) finding E as a function of V (that is, $V_{ia\sigma}$ and $V_{ic\sigma}$, where $i=s$, p_x, p_y, or p_z) by solving the secular Eq. (4) using the known Hamiltonian H^0, and (ii) determining the appropriate self-consistent potential V, and the charge distribution determining it. Since (i) has been discussed in detail elsewhere,[1,2] we explain only (ii).

For a free atom, the one-electron energy of the valence electron in spin-orbital α ($\alpha=s\uparrow$, $s\downarrow$, $p_x\uparrow$, $p_y\uparrow$, $p_z\uparrow$, $p_x\downarrow$, $p_y\downarrow$, or $p_z\downarrow$) for the given configuration $\{n_\alpha\}$ depends on the configuration and is approximated by the following expression of Haldane and Anderson,[10] using three different electron repulsion parameters U_{ss}, U_{pp}, and U_{sp}:

$$E_{s\sigma}\{n_\alpha\}=E_s^0+\sum_{\sigma'}{}'n_{s\sigma'}U_{ss}+\sum_{j,\sigma'}n_{p_j\sigma'}U_{sp} , \tag{7}$$

and

$$E_{p_j\sigma}\{n_\alpha\}=E_p^0+\sum_{j,\sigma'}{}'n_{p_j\sigma'}U_{pp}+\sum_{\sigma'}n_{s\sigma'}U_{sp}\ , \tag{8}$$

where σ is the spin ($\uparrow$ or $\downarrow$), we have $j=x$, y, or z, and the prime on the summation indicates that the self-interaction is excluded. Here n_α are the occupation numbers of spin-orbital α and they are integers (0 or 1) for the free atom. Sankey and Dow[11] determined the five empirical parameters E_s^0, E_p^0, U_{ss}, U_{pp}, and U_{sp}, using the requirement that Hartree-Fock s- and p-electron energies and the observed ionization potentials of the free atoms be fitted. In the solid, we assume that the electronic energy of an atom is a continuous and differentiable function of occupation numbers[14] n_μ with the same empirical parameters U as those of the free atom— but with n_μ not necessarily integers. Here we have the notation $\mu\equiv(\alpha,b,\vec{\mathrm{R}})$, where b is either a (for anion) or c (for cation) and $\vec{\mathrm{R}}$ denotes the position of the unit cell in the crystal. The anion site of the central cell [see Eq. (6)] is denoted by $\vec{\mathrm{D}}=(a,\vec{0})$.

The spin-orbital occupation number is $n_\mu=n_\mu^{\mathrm{deep}}+n_\mu^v$, where n_μ^v comes from the redistributed electrons in the valence bands and can be found by integrating the local spectral density of states $D_\mu(E)$ from $-\infty$ to the top of the valence band (zero energy), i.e.,

$$n_\mu^v=-\int_{-\infty}^{0}f(E)D_\mu(E)dE\ , \tag{9}$$

where $f(E)$ is unity if the one-electron state of energy E is occupied and zero otherwise. The spectral density of states $D_\mu(E)$ is related to the Green's operator $G^0(E)$ of the perfect crystal by[15]

$$\begin{aligned}D_\mu(E)&=\langle\mu\,|\,\rho(E)\,|\,\mu\rangle=(-1/\pi)\mathrm{Im}\langle\mu\,|\,G(E)\,|\,\mu\rangle\ ,\\&=(-1/\pi)\mathrm{Im}\langle\mu\,|\,[1-G^0(E)V]^{-1}G^0(E)\,|\,\mu\rangle\ ,\end{aligned} \tag{10}$$

where $\rho(E)=(-1/\pi)\mathrm{Im}G(E)$ is the state density operator for the perturbed crystal, and the last relation comes from Dyson's equation.

The total spin-orbital occupation number includes a contribution from the deep level and is given by

$$n_\mu=n_\mu^v+\sum_i\langle\mu\,|\,\psi_i\rangle\langle\psi_i\,|\,\mu\rangle\ , \tag{11}$$

where $|\,\psi_i\rangle$ are the wave functions of occupied discrete states in the band gap. The wave function of the discrete state at energy E in the band gap can be obtained by solving Eq. (3) with the normalization condition[15]

$$\begin{aligned}1&=\langle\psi\,|\,\psi\rangle=\langle\psi VG^0(E)\,|\,G^0(E)V\psi\rangle\\&=\langle\psi\,|\,V[G^0(E)]^2V\,|\,\psi\rangle\ .\end{aligned} \tag{12}$$

Since we have

$$[G^0(E)]^2=-\left[\frac{d}{dE}\right]G^0(E)\ , \tag{13}$$

the normalization condition becomes

$$\langle\psi\,|\,V\left[\frac{d}{dE}\right]G^0(E)V\,|\,\psi\rangle=-1\ . \tag{14}$$

Finally, the new matrix elements of the impurity potential can be constructed by the following:

$$\begin{aligned}V_\mu&=\langle\mu\,|\,V\,|\,\mu\rangle=\langle\mu\,|\,H-H^0\,|\,\mu\rangle\\&=E_\mu(\mathrm{impurity})-E_\mu(\mathrm{host})\ .\end{aligned} \tag{15}$$

Now, the self-consistent scheme is implemented as follows: For the input impurity potential V_μ on each site, we solve Eq. (4) for the eigenvalue E. With this eigenvalue E and input V_μ, we compute the total spin-orbital occupation numbers n_μ for the site by Eq. (11). These n_μ give the new V_μ by Eqs. (15), (7), and (8). This procedure is repeated iteratively until self-consistency is obtained.

III. RESULTS AND DISCUSSION

The above method was applied to S^+, S^0, Se^+, Se^0, Te^+, and Te^0 donors in Si, because they are well-studied substitutional impurities and experimental ionization energies are available. For the unperturbed host Si band structure, we used the empirical tight-binding model of Vogl *et al.*,[9] which yields good band structures, including the lowest conduction band, with an indirect band gap of 1.17 eV. In computing the defect levels, the self-consistency scheme was iterated for those μ referring to the central cell only.[16] Then, using Eq. (3), the wave functions at the first-neighbor sites were computed, as well as the charge densities n_μ [Eq. (9)] and the defect potentials [Eq. (15)] for μ referring to these sites. Then Eq. (4) was solved for the defect level (without iterating the defect potentials on the neighboring sites to self-consistency).[16]

Our calculations show that the neutral and singly ionized centers each form an s-like A_1 state in the band gap and a triply degenerate p-like T_2 resonance state just above the conduction-band edge. This A_1 state is pulled down from the conduction band because the chalcogenides are more electronegative than Si. The A_1 state is occupied by one and two electrons for the singly ionized and neutral centers, respectively. Although the charge states of the two levels differ by unity, only about 8% of each deep-level electron's charge resides within the central cell of the impurity.[17–19] Ionization of the neutral impurity decreases $\sum_\alpha n_{\alpha\,\vec{\mathrm{D}}}^{\mathrm{deep}}$ from 0.16 to 0.08, but $n_{\alpha\,\vec{\mathrm{D}}}^{v}$ increases to almost fully compensate this effect. Thus the valence electrons screen the deep impurity to make it locally neutral in the central cell, regardless of its global charge state.

The predicted absolute ionization energies of the chalcogen's deep levels are given in Fig. 1, where they are compared with the data of Refs. 3 and 4. The agreement is gratifying, especially since the theory omits the effects of lattice relaxation,[20] the long-ranged Coulombic electron-impurity interaction, and electron-electron correlations—and hence can be expected to have an uncertainty of a few tenths of an eV. Indeed, the agreement between theory and data becomes excellent if the theory is shifted downward by $\simeq$0.3 eV. We are aware of one other self-consistent calculation of a charge-state splitting for a chalcogen substitutional impurity in Si: Bernholc *et al.*[8] treated S in Si. That theory predicted an A_1 state 0.1 eV

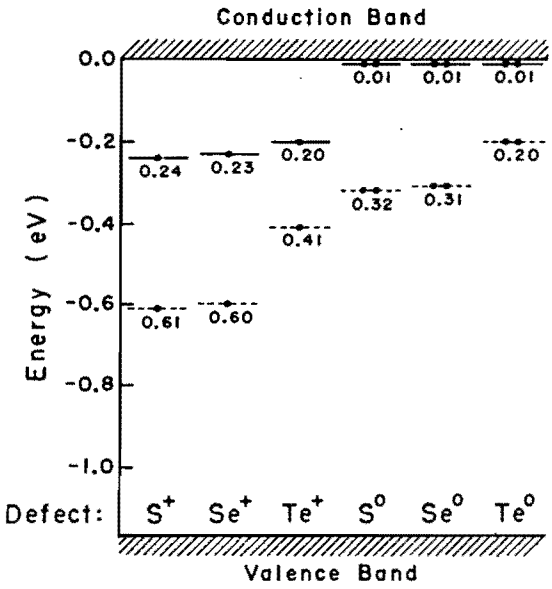


FIG. 1. Deep energy levels in the band gap of Si of the singly charged and neutral chalcogen impurities. All energies are measured with respect to the conduction-band minimum or ionization threshold. The theoretical predictions are denoted by solid lines (with their energies underneath) and the data of Refs. 3 and 4 are given by dashed lines. Occupation of the A_1 one-electron states by one or two electrons is denoted by solid circles.

below the conduction-band minimum, and found T_2- and E-symmetric resonances somewhat above the band edge. After including the long-ranged Coulombic electron-impurity interaction, Bernholc *et al.* predicted first and second ionization energies for S of 0.25 and 0.45 eV—in satisfactory agreement with our results of 0.01 and 0.24 eV. The theoretical uncertainties in the self-consistent local density calculations are comparable with those of the present work despite the orders of magnitude greater computational complexity of that theory, which, because of problems related to obtaining the correct energy-band gap, must include many bands to obtain an adequate representation of the conduction-band spectral density in Eq. (4). (The present work circumvents that problem, in effect, by using an empirical fit to the conduction-band structure and the density of states.[9]) Shimizu and Minami, using a cluster molecular-orbital theory, predicted an ionization energy of S^+ in Si of 0.57 eV.[21] While this energy is in excellent agreement with the data, the wave function associated with this level appears to be quite different from that determined experimentally.[18,22]

The charge-state splittings of ionization energies predicted by the present theory are 0.23, 0.22, and 0.19 eV, and are in good agreement with the experimental values 0.30, 0.29, and 0.21 eV (Refs. 3 and 4) for S, Se, and Te, respectively. The chemical trend in the observed charge-state splittings, $\Delta E_S > \Delta E_{Se} > \Delta E_{Te}$, is correctly reproduced. The value of 0.20 eV obtained for S by Bernholc *et al.*[8] is slightly farther from the data than our value, but this difference is not significant, and the two theories should be viewed as giving the same prediction.

The predictions of deep-level energies obtained here are very similar to those predicted by Hjalmarson *et al.*[1,2] using a non-self-consistent theory for neutral impurities. By iterating the theory to self-consistency we have been able to obtain charge-state splittings of the magnitude observed experimentally. However, the present work shows that a deep impurity tends to remain locally neutral in its central cell, regardless of its global charge state. For deep levels associated with single or double donors or acceptors in covalent homopolar semiconductors, the splittings are known to be 0.2 to 0.3 eV in magnitude, and can just as accurately be taken into account by an *ad hoc* adjustment upwards (downwards) of the neutral defect levels by $\simeq 0.2$ to 0.3 eV for each extra electron (hole). Of course, this should not apply to highly charged states of defects in strongly heteropolar materials or to systems in which there is significant charge transfer. For such systems, fully self-consistent theories may be necessary.

ACKNOWLEDGMENTS

We are grateful to the U.S. Army Research Office for their support of this work (Contract No. DAAG29-83-K-0122). We thank K. E. Newman for her stimulating comments and for her constructive reading of the manuscript.

[1]H. P. Hjalmarson, Ph. D. thesis, University of Illinois, Urbana, 1979.
[2]H. P. Hjalmarson, P. Vogl, D. J. Wolford, and J. D. Dow, Phys. Rev. Lett. 44, 810 (1980), and unpublished.
[3]H. G. Grimmeiss, E. Janzén, and K. Larsson, Phys. Rev. B 25, 2627 (1982).
[4]H. G. Grimmeiss and B. Skarstam, Phys. Rev. B 23, 1947 (1981).
[5]See, for example, L. I. Schiff, *Quantum Mechanics*, 3rd ed. (McGraw-Hill, New York, 1968), p. 259.
[6]C. E. Moore, *Atomic Energy Levels as Derived from the Analyses of Optical Spectra*, National Bureau of Standards Circular No. 467 (U.S. G.P.O., Washington, D. C., 1949, 1952, and 1958), Vols. I–III.
[7]H. G. Grimmeiss, E. Janzén, H. Ennen, O. Schirmer, J. Schneider, R. Wörner, C. Holm, E. Sirtl, and P. Wagner, Phys. Rev. B 24, 4571 (1981).
[8]J. Bernholc, N. O. Lipari, S. T. Pantelides, and M. Scheffler, Phys. Rev. B 26, 5706 (1982).
[9]P. Vogl, H. P. Hjalmarson, and J. D. Dow, J. Phys. Chem. Solids 44, 365 (1983).
[10]F. D. M. Haldane and P. W. Anderson, Phys. Rev. B 13, 2553 (1976).
[11]O. F. Sankey and J. D. Dow, Phys. Rev. B 27, 7641 (1983).
[12]O. F. Sankey and J. D. Dow, J. Appl. Phys. 52, 5139 (1981).
[13]W. A. Harrison, *Electronic Structure and the Properties of Solids* (Freeman, San Francisco, 1980).
[14]J. C. Slater, *Quantum Theory of Molecules and Solids* (McGraw-Hill, New York, 1974), Vol. 4, p. 43.
[15]J. Bernholc and S. T. Pantelides, Phys. Rev. B 18, 1780 (1978); J. Bernholc, N. O. Lipari, and S. T. Pantelides, Phys. Rev. B 21, 3545 (1980).

[16]Self-consistency was considered to have been obtained when the defect potential did not change by 0.001 eV from one iteration to the next. We did carry the calculations to self-consistency in the first-neighbor potentials for the neutral impurities S^0, Se^0, and Te^0, and found that the absolute energies of the deep levels changed by a value only $\simeq 0.05$ eV, a negligible amount on the scale of theoretical uncertainty.

[17]The value of 8% obtained here differs somewhat from the value obtained in non-self-consistent calculations of Refs. 18 and 19, because the energies of the levels differ somewhat.

[18]S. Y. Ren, W. M Hu, O. F. Sankey, and J. D. Dow, Phys. Rev. B 26, 951 (1982).

[19]O. F. Sankey and J. D. Dow, Solid State Commun. 51, 705 (1984).

[20]M. Scheffler, J. P. Vigneron, and G. B. Bachelet, Phys. Rev. Lett. 49, 1765 (1982).

[21]T. Shimizu and K. Minami, Phys. Status Solidi B 48, K181 (1971).

[22]G. W. Ludwig, Phys. Rev. 137, A1520 (1965).

Localized Perturbations in Semiconductors.

JOHN D. DOW

Department of Physics, University of Notre Dame - Notre Dame, IN 46556

1. – Introduction.

In these lectures we outline a simple but general theory of electronic states associated with localized perturbations in semiconductors. The basic problem we consider is the «deep-level problem», namely predicting the point-defect energy levels that lie near the middle of the band gap of a semiconductor. When we began work on this problem, a deep level was defined as a level that was not shallow, namely one more than 0.1 eV from the nearest band edge—a level that could not be thermally ionized at room temperature. (That definition has since been revised: see below.) Our own interest in the deep-level problem resulted from data of Wolford and Streetman for the N impurity in $GaAs_{1-x}P_x$ alloys [1]. This impurity appeared to be shallow in GaP, having a binding energy of only $\simeq 11$ meV, even smaller than the 35 meV effective-mass theory binding energy of the shallow donors S and Se. However, it became a genuine deep level in the alloy for $x \simeq 0.5$ and merged into the conduction band as a resonance for $x < 0.22$ (see fig. 1 [2, 3]). Thus the N impurity level was apparently shallow (for $x = 1$), deep (for $x \simeq 0.5$) and no level at all (for $x < 0.22$) as one varied alloy composition x continuously from GaP ($x = 1$) to GaAs ($x = 0$).

1·1. $GaAs_{1-x}P_x$. – The alloy host $GaAs_{1-x}P_x$ has a band structure that is well described by the virtual-crystal approximation [4] and varies continuously from the direct-gap band structure of GaAs (with the conduction band minimum at $\Gamma = (0, 0, 0)$ in the Brillouin zone) to the indirect-gap structure of GaP (with the conduction band minimum near the X-point: $(2\pi/a_L)(1, 0, 0)$) (see fig. 2). The band gap of GaAs is in the infra-red. Pure GaAs would emit such light because the band gap is direct, and the magnitude of the momentum of a thermalized electron-hole pair, $|\boldsymbol{k}_e - \boldsymbol{k}_h|$, can be equal to that of the emitted photon, $2\pi/\lambda$, which is essentially zero on the scale of the Brillouin zone. In contrast, GaP has an indirect-gap band structure, and so a thermalized conduction electron has a significantly different wave vector from a thermalized

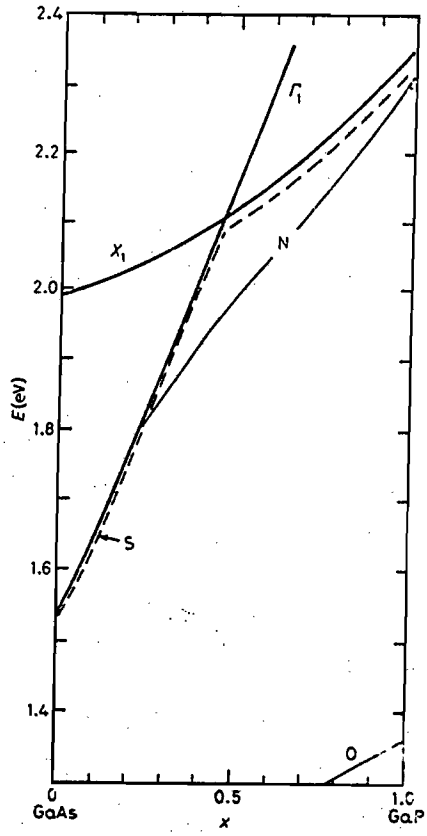


Fig. 1. – Illustration of the dependences of shallow and deep impurity levels on alloy composition x in $GaAs_{1-x}P_x$ alloys, after ref. [2, 3]. The zero of energy is the valence band maximum. The direct conduction band edge is Γ_1 and the indirect edge is X_1. The N and O deep levels are denoted by solid lines. The shallow levels of S (or Se) are denoted by dashed lines. Note that the direct-indirect cross-over occurs for $x \simeq 0.45$ and that the shallow-level binding energy is larger in indirect material (because the effective mass is larger).

hole. Hence pure GaP cannot emit light even though its band gap is in the green—a highly visible part of the spectrum. The alloy has become technologically important, because, for $x \simeq 0.4$, the band gap lies in the visible (red), but the band structure is still direct—hence this material is employed in red light-emitting diodes (LEDs).

1·2. *Column V site impurities* N, O, S *and* Se *in* $GaAs_{1-x}P_x$. – To fabricate a light-emitting diode that emits in the yellow or the green from these alloys, one needs a source or sink of crystal momentum, $\boldsymbol{K} \simeq (2\pi/a_L)(1, 0, 0)$, so that the selection rule $\boldsymbol{k}_e - \boldsymbol{k}_h \pm \boldsymbol{K} \simeq 0$ can be satisfied. Impurities can supply the

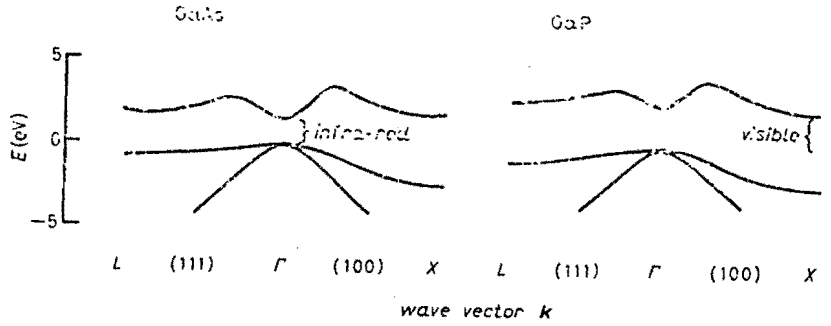


Fig. 2. – Electronic energy band structures $E(k)$ of GaAs and GaP from L to Γ to X along the (100) (*i.e.* Γ to X) and (111) (*i.e.* Γ to L) directions of the Brillouin zone, after M. L. COHEN and T. K. BERGSTRESSER: *Phys. Rev.*, **141**, 789 (1966). Note that the band gap of GaAs is direct but in the infra-red; GaP has an indirect gap from the valence band maximum at Γ to the conduction band minimum at X, in the visible region of the spectrum.

needed momentum, with the impurities most likely to occur on the column V site of $GaAs_{1-x}P_x$ being N, O, S and Se. Ironically, two of these impurities, S and Se, produce *shallow* levels in the band gap of $GaAs_{1-x}P_x$ that lie close to the conduction band edge and follow the edge as the composition varies. But two do not. Oxygen lies several tenths of an electronvolt deep in the band gap of GaP and its energy level decreases linearly as x decreases—it is a genuine deep level by all definitions. The behavior of N (with respect to the valence band maximum) is especially interesting: in GaP it is *apparently shallow* with a 11 meV binding energy, and, with decreasing alloy composition x, its energy level decreases linearly, similar to the oxygen deep level, becoming a genuine deep trap (by the old definition: more than 0.1 eV from the conduction band edge) for $x \simeq 0.5$. At $x = 0.22$, however, the N level goes into the conduction band. In other words, N appears to be shallow energetically for $x = 1$, is deep for $x = 0.5$ and is a resonance for $x = 0$ (see fig. 1). Moreover, the N level is unattached to the conduction band edge, and dE/dx for this level is characteristic of a deep trap such as O. These facts led us to believe that N is, in fact, a deep level whose energy accidentally lies close to the conduction band edge in GaP and becomes resonant in GaAs—and focussed our attention on N as the prototypical deep trap [5].

2. – The Vogl model of electronic structure.

The foundation for much of what we shall discuss in these lectures is an empirical tight-binding theory of electronic structure developed by VOGL *et al.* [6]. This theory has three distinguishing features: i) It properly represents

the chemistry of the sp^3 bonding, because it has a basis which includes one s orbital and three p orbitals at each atomic site (as well as one additional s orbital: s^*). ii) It produces indirect-gap band structures for Si and GaP with a minimum number of basis functions, five per site. (Pure sp^3-basis tight-binding models do not; the extra s^* orbital produces the indirect band structure by pushing the indirect conduction band minimum down in energy). iii) The parameters of the Hamiltonian exhibit manifest chemical trends that are codified in scaling rules: the diagonal matrix elements are related to atomic energies, and the off-diagonal matrix elements are inversely proportional to d^2, the square of the bond length (Harrison's rule [7]).

This empirical Hamiltonian was arrived at by VOGL and HJALMARSON after a great deal of labor and represents an attempt to simultaneously describe the energy bands of sixteen semiconductors. The Vogl model drew much of its inspiration from Harrison's bond orbital model [8], which was one of the first successful attempts to develop a simple Hamiltonian for describing chemical trends for many semiconductors—it described valence band structures rather accurately. A distinguishing feature of the Vogl model is its ability to reproduce general features of the lowest conduction bands as well.

The scaling rules for chemical trends in the parameters of the Vogl model are very important. Because of them, the Vogl Hamiltonian can be generalized to treat *inhomogeneous* semiconductors—even though the information contained in the model's parameters comes exclusively from the known energy band structures of homogeneous semiconductors. For example, if one atom is changed (*e.g.*, one P in GaP is replaced by a N atom), the matrix elements for the changed Hamiltonian can be deduced by changing the host matrix elements according to the scaling rules.

The basic philosophy in real ($\boldsymbol{R}$) space of empirical tight-binding theory is similar to the philosophy for ordinary pseudopotential theory (in $\boldsymbol{k}$-space): remove the distant parts (in $\boldsymbol{R}$-space) of the Hamiltonian and lump them into near-neighbor parameters that are determined empirically.

Tight-binding basis functions $|nb\boldsymbol{k}\rangle$ are contructed from (unknown) localized quasi-atomic orbitals $|nb\boldsymbol{R}_j\rangle$

$$|nb\boldsymbol{k}\rangle = N^{-\frac{1}{2}} \sum_j |nb\boldsymbol{R}_j\rangle \exp\left[i\boldsymbol{k}\cdot\boldsymbol{R}_j + i\boldsymbol{k}\cdot\boldsymbol{v}_b\right], \tag{1}$$

where $n = s, p_x, p_y, p_z$, or s^* specifies the basis orbital, we have $b =$ a (anion) or $b =$ c (cation), $\boldsymbol{k}$ is the Bloch wave vector, $\boldsymbol{R}_j$ is an anion site in a zincblende structure, we have $\boldsymbol{v}_a = \boldsymbol{0}$, and $\boldsymbol{v}_c$ is the position of the cation relative to the anion. In this basis, the secular equation reduces to the 10×10 system

$$(H^0 - \varepsilon(k\lambda))|k\lambda\rangle = 0, \tag{2}$$

where we have (symbolically) the Bloch state

$$|k\lambda\rangle = \sum_{n,b} |nbk)(nbk|k\lambda\rangle \,. \tag{3}$$

In the $|nbk)$ tight-binding basis we have the 10×10 Hamiltonian matrix ($\lambda = 1, 2, \ldots, 10$)

(4)

	$\|sa)$	$\|sc)$	$\|p_x a)$	$\|p_y a)$	$\|p_z a)$	$\|p_x c)$	$\|p_y c)$	$\|p_z c)$	$\|s^* a)$	$\|s^* c)$
$(sa\|$	$E(s,a)$	$V(s,s)g_0$	0	0	0	$V(sa,pc)g_1$	$V(sa,pc)g_2$	$V(sa,pc)g_3$	0	0
$(sc\|$	$V(s,s)g_0^*$	$E(s,c)$	$-V(pa,sc)g_1^*$	$-V(pa,sc)g_2^*$	$-V(pa,sc)g_3^*$	0	0	0	0	0
$(p_x a\|$	0	$-V(pa,sc)g_1$	$E(p,a)$	0	0	$V(x,x)g_0$	$V(x,y)g_3$	$V(x,y)g_2$	0	$-V(pa,s^*c)g_1$
$(p_y a\|$	0	$-V(pa,sc)g_2$	0	$E(p,a)$	0	$V(x,y)g_3$	$V(x,x)g_0$	$V(x,y)g_1$	0	$-V(pa,s^*c)g_2$
$(p_z a\|$	0	$-V(pa,sc)g_3$	0	0	$E(p,a)$	$V(x,y)g_2$	$V(x,y)g_1$	$V(x,x)g_0$	0	$-V(pa,s^*c)g_3$
$(p_x c\|$	$V(sa,pc)g_1^*$	0	$V(x,x)g_0^*$	$V(x,y)g_3^*$	$V(x,y)g_2^*$	$E(p,c)$	0	0	$V(s^*a,pc)g_1^*$	0
$(p_y c\|$	$V(sa,pc)g_2^*$	0	$V(x,y)g_3^*$	$V(x,x)g_0^*$	$V(x,y)g_1^*$	0	$E(p,c)$	0	$V(s^*a,pc)g_2^*$	0
$(p_z c\|$	$V(sa,pc)g_3^*$	0	$V(x,y)g_2^*$	$V(x,y)g_1^*$	$V(x,x)g_0^*$	0	0	$E(p,c)$	$V(s^*a,pc)g_3^*$	0
$(s^* a\|$	0	0	0	0	0	$V(s^*a,pc)g_1$	$V(s^*a,pc)g_2$	$V(s^*a,pc)g_3$	$E(s^*,a)$	$V(s^*,s^*)g_0$
$(s^* c\|$	0	0	$-V(pa,s^*c)g_1^*$	$-V(pa,s^*c)g_2^*$	$-V(pa,s^*c)g_3^*$	0	0	0	$V(s^*,s^*)g_0^*$	$E(s^*,c)$

Here we have for $k = (2\pi/a_L)(k_1, k_2, k_3)$ and a_L the lattice constant

$$g_0(k) = \cos(\pi k_1/2)\cos(\pi k_2/2)\cos(\pi k_3/2) - i\sin(\pi k_1/2)\sin(\pi k_2/2)\sin(\pi k_3/2)\,,$$

$$g_1(k) = -\cos(\pi k_1/2)\sin(\pi k_2/2)\sin(\pi k_3/2) + i\sin(\pi k_1/2)\cos(\pi k_2/2)\cos(\pi k_3/2)\,,$$

$$g_2(k) = -\sin(\pi k_1/2)\cos(\pi k_2/2)\sin(\pi k_3/2) + i\cos(\pi k_1/2)\sin(\pi k_2/2)\cos(\pi k_3/2)$$

and

$$g_3(k) = -\sin(\pi k_1/2)\sin(\pi k_2/2)\cos(\pi k_3/2) + i\cos(\pi k_1/2)\cos(\pi k_2/2)\sin(\pi k_3/2)\,.$$

We have

(5)

$$\begin{aligned}
&(saR|H|saR) = E(s,a)\,,\\
&(paR|H|paR) = E(p,a)\,,\\
&(scR|H|scR) = E(s,c)\,,\\
&(pcR|H|pcR) = E(p,c)\,,\\
&(s^*aR|H|a^*aR) = E(s^*,a)\,,\\
&(s^*cR|H|s^*cR) = E(s^*,c)\,,\\
&4(saR|H|scR) = V(s,s)\,,\\
&4(p_x aR|H|p_x cR) = V(x,x)\,,\\
&4(p_x aR|H|p_y cR) = V(x,y)\,,\\
&4(saR|H|p_x cR) = V(sa,pc)\,,\\
&4(p_x aR|H|scR) = V(sc,pa)\,,\\
&4(s^*aR|H|p_x cR) = V(s^*a,pc)\\
&\text{and}\\
&4(p_x aR|H|s^*cR) = V(pz,s^*c)\,.
\end{aligned}$$

This is the basic Hamiltonian to be used throughout the present work. The reader should become familiar with it by working the following three problems.

Problem 1. Compute the energy band structure at $\boldsymbol{k} = 0$ of GaP, taking your zero of energy at the valence band maximum. Compare your results with fig. 2 of ref. [6]. For GaP the Vogl tight-binding matrix elements are $E(s, a) = -8.1124$, $E(p, a) = 1.1250$, $E(s, c) = -2.1976$, $E(p, c) = 4.1150$, $E(s^*, a) = 8.5150$, $E(s^*, c) = 7.1850$, $V(s, s) = -7.4709$, $V(x, x) = 2.1516$, $V(x, y) = 5.1369$, $V(sa, pc) = 4.2771$, $V(sc, pa) = 6.3190$, $V(s^*a, pc) = 4.6541$ and $V(pa, s^*c) = 5.0950$. (For other semiconductors, see ref. [6].)

Problem 2. Compute the band structure at the X-point, $\boldsymbol{k} = (2\pi/a_L)(1, 0, 0)$, of the Brillouin zone.

Problem 3. Write down the change in the Hamiltonian matrix, ΔH, in the $|nb\boldsymbol{R})$ basis for a N atom replacing P in GaP at $\boldsymbol{R} = \boldsymbol{D}$. Assume that the bond length does not change when N replaces P and that the matrix elements involving s^* remain unaltered (because s^* simulates nonlocal effects of distant neighbors). Neglect distinctions between the host basis orbitals $|nb\boldsymbol{R})$ and the corresponding impurity orbitals (in subsequent work, we shall actually be using the impurity orbitals at the impurity site). Show that the matrix is 4×4 and diagonal. Suppose further that the diagonal matrix elements V_s and V_p of ΔH are given by the Vogl-Hjalmarson scaling rules [6]

$$V_s = 0.8\big(w(s, \mathrm{N}) - w(s, \mathrm{P})\big) \tag{6a}$$

and

$$V_p = 0.6\big(w(p, \mathrm{N}) - w(p, \mathrm{P})\big)\,, \tag{6b}$$

where the atomic-orbital energies w for N and P are [6] $w(s, \mathrm{N}) = -25.7130$, $w(p, \mathrm{N}) = -15.4388$, $w(s, \mathrm{P}) = -18.9425$ and $w(p, \mathrm{P}) = -10.6544$. Finally, using ref. [6], determine the numerical values of the defect potentials V_s and V_p for O, S, Se and Te and for B, C, N, O and F. If you have worked problem 3, you have set up the Hamiltonian for obtaining the deep levels of N in GaP.

3. – The Hjalmarson *et al.* theory of deep impurity levels.

3·1. *Qualitative remarks.* – In the late 1950's KOHN and collaborators developed the effective-mass theory of shallow impurities in semiconductors [9]. According to this theory, an impurity such as S substituting for As in GaAs produces a donor electron that orbits the extra nuclear charge of S (relative

to As) in a large hydrogenic orbit, the envelope wave function for which satisfies a Schrödinger equation

$$[(-\hbar^2/2m^*)\nabla - (Ze^2/\varepsilon r)]\psi = (E - E_c)\psi\,, \tag{7}$$

where m^* is the conduction band effective mass, Z is the excess valence of the impurity atom with respect to the host atom it replaces (unity for S on an As site in GaAs), ε is the GaAs static dielectric constant, and E_c is the energy of the conduction band edge (at $k = 0$ in GaAs). This effective-mass state has a total wave function that is primarily a product of this envelope function and the periodic part of a Bloch function evaluated at the wave vector of the conduction band minimum [9, 10]: it is made up primarily from one band (the GaAs conduction band, in this case). The effective-mass state is hydrogenic and virtually 100 % hostlike. The impurity level is « attached » to the conduction band edge with a small binding energy of order 10 meV (13.6 eV$(m^*/m_0)^2/\varepsilon$, where m_0 is the free-electron mass) and follows the edge when the edge moves as a result of externally applied pressure or alloying (*e.g.*, alloying GaAs with GaP). The shallow levels control the electrical properties of the semiconductor, and, although the impurity potential in the central cell often deviates greatly (a few eV) from the Coulombic value, $-Ze^2/\varepsilon r$, only the long-ranged Coulombic potential seems to have a significant effect on the shallow states. (This should be bothersome, because central-cell potentials of order 1 eV must produce some effect on that scale.) Moreover, the shallow states are localized in k-space but delocalized in real space.

The effective-mass theory accounts for many of the data for impurity levels in the band gaps of semiconductors; however, it does not account for many facts, including the following: i) some isoelectronic impurities, such as N replacing P in GaP, produce levels in the gap despite the fact that their valence differences Z are zero, and ii) some levels in the gap lie far (more than 0.1 eV) from a band edge and are « deep levels ». Early attempts to explain these facts attempted to modify the effective-mass theory to produce larger binding energies.

A central point of the Hjalmarson theory [11, 12] is that *every* heterovalent substitutional impurity produces *both* « deep levels » and shallow levels, and that *the « deep levels » do not necessarily lie in the fundamental band gap, but may be resonant with the host bands.* The deep and shallow states are two qualitatively different types of impurity states that coexist, but are rarely observed simultaneously. Deep levels are controlled by the central-cell potential, have wave functions that are linear combinations of wave functions from many ($\simeq 10$) host bands, are often antibonding in character and are largely hostlike. The deep-level energies are often unattached to nearby band edges and do not follow them when they move as a result of pressure or alloying. Deep impurity states are localized in real space and delocalized in k-space. The apparent

« binding energy » of a deep level relative to a nearby band edge is often large (tenths of eV) in magnitude and can be negative. When a deep level falls in the fundamental band gap, it can trap excitons or charge carriers, often enhancing the nonradiative recombination of electrons and holes. Thus deep levels tend to influence the optical properties of semiconductors even at concentrations as low as $10^{15}/cm^3$.

In these lectures, we limit ourselves to levels associated with sp^3-bonded substitutional impurities. (For discussions of interstitial sp^3-bonded impurities and transition metal impurities, see ref. [13] and [14], respectively.) Thus, in the energy vicinity of the band gap, for substitutional impurities in tetrahedral semiconductors, we expect exactly four deep levels to originate from the sp^3 bonding, three of which are degenerate: a s-like A_1 level and a p-like triply degenerate T_2 level. (A_1 and T_2 are irreducible representations of the tetrahedral group T_d.) If these « deep levels » due to the central-cell potential all happen to lie above the conduction band edge, as in the case of $GaP:S_P$ (S on a P site in GaP), then the only levels in the gap are the shallow levels associated with the long-ranged Coulomb potential—and S is termed a shallow impurity because only its shallow levels are observed in the gap. If one of the deep levels due to the central-cell potential falls within the fundamental band gap, as with $GaP:O_P$, then the impurity is termed « deep ». But a central point is that both shallow and deep levels of the same impurity coexist (fig. 3); they are distinct (although deep levels near a band edge may hybridize with shallow levels). Isoelectronic defects, such as $GaP:N_P$, have no long-ranged Coulomb potential and hence no shallow levels; all of their defect levels (except possibly levels associated with a strain field surrounding them) are « deep ».

3·2. *Energy scales and the nature of the theory.* – Before constructing a theory of deep impurity levels, one should first determine the important physics.

To begin with, the bonding in semiconductors is sp^3 in character, and a proper treatment of a localized defect state must account for this. The spectral distribution of the sp^3 bonds covers $\simeq 20$ eV, the combined widths of the valence bands and the lowest conduction bands.

The defect potential in the central cell can be crudely estimated as the difference between the atomic energies of the defect and the host atom it replaces —and is typically *several electronvolt* in magnitude—of order 5 eV, 7 eV and 15 eV for S, N and O (all on the P site) in GaP, and 4 eV for P in Si. The fact that the central-cell defect potential is so large should be extremely puzzling, especially in the case of the shallow donors P in Si and S_P in GaP—because Nature requires that a perturbation of several electronvolt exhibit itself on a scale of order of electronvolt, and the shallow impurities appear at first glance to exhibit consequences of the central-cell defect potential on only the millielectronvolt scale. The resolution of this dilemma lies in the fact that the shallow donors also produce « deep resonances », quasi-localized states at energies of order 1 eV

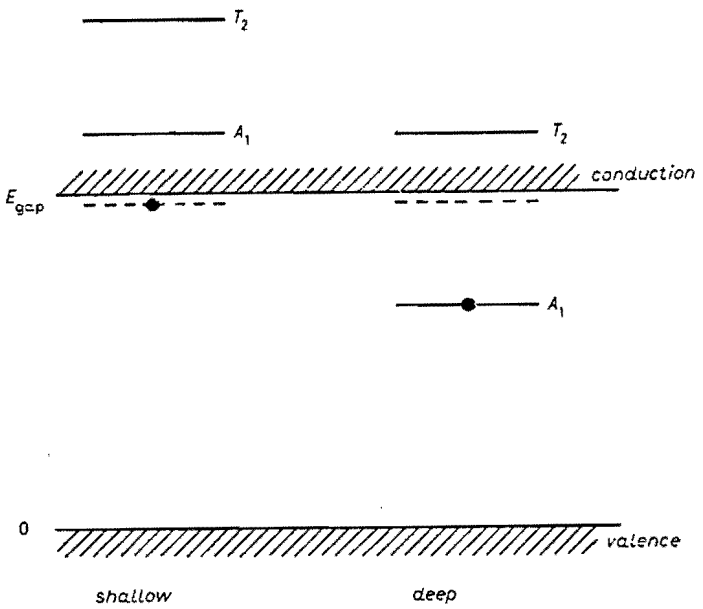


Fig. 3. – Schematic illustration of the difference between « shallow » and « deep » sp^3-bonded substitutional (donor) impurities, after ref. [3]. The shallow energy levels in the band gap are dashed. The deep levels of A_1 (s-like) and T_2 (p-like) symmetry are denoted by heavy lines. In the case of a « shallow impurity » the deep levels are resonances and lie outside the fundamental band gap; for a « deep impurity » at least one deep level lies within the gap. The lowest level is occupied by the extra electron (dark circles) if the impurity has a valence one greater than the host atom it replaces (*e.g.*, S or O on a P site in GaP).

above the conduction band minimum; the existence of these resonances has been appreciated only recently. This notion, that impurities produce « deep » levels above the conduction band minimum, requires a new definition of « deep ». The old definition was that any level in the fundamental band gap more than 0.1 eV from the nearest band edge was « deep ». Now, following HJALMARSON *et al.*, we *define* a deep level as one whose physics is controlled by the central-cell potential; as a result, « deep » levels now may have very small (< 0.1 eV) binding energies (such as the N level in GaP) or may lie above the conduction band edge with « negative binding energies », resonant with the host bands (« deep resonances »). They may also lie resonant with the valence bands (binding energies greater than the band gap).

In addition to the central-cell potential, several other physical effects influence deep levels on a scale of a few tenths of an eV. These include lattice relaxation around the defect [15] and charge state splitting [16, 17] of the defect levels (*e.g.*, the difference due to electron-electron interactions in the S^0 and S^- one-electron energy levels in Si). Furthermore, the Coulomb potential outside the central cell, $-Ze^2/\varepsilon r$, is also of order 0.1 eV for $r > a_L$. In much of the work we

discuss, we neglect all such effects of order 0.1 eV and concentrate on predicting the *chemical trends* in deep energy levels, from one impurity to another or from one host to another. By neglecting these effects, we obtain a very simple theory in which the defect potential matrix is diagonal (in a localized basis) and is localized to the central cell of the defect—because of the scaling rules of the Vogl Hamiltonian.

The band gap of a typical semiconductor is $\simeq 1$ eV and studies of deep levels are directed toward determining the levels with an accuracy of < 0.1 eV. However, the band gap energy is *not* a scale of physical relevance to the deep-impurity problem, because deep levels (whether « bound » in the gap or resonant) are unattached to band edges. Instead the band gap energy determines the scale of *observability* of most deep levels. A simple way to think about deep impurity levels is that they lie throughout the $\simeq 20$ eV range of the sp^3 bond, but that only the small fraction of these levels that lies within the « window » of the band gap is observable by conventional means. Hence a complete description of deep-level experiments on the scale of observability of deep levels requires a theory with an accuracy of 0.1 eV out of 20 eV, or 0.5 %. No contemporary theory is capable of such an accuracy; the best accuracy achievable is a few tenths of an eV. Therefore, the goal of theory should not be to predict the absolute energies of deep levels in the band gap, because this goal is presently unattainable. Rather, theories should be constructed with the intent of simply displaying the physics of deep levels, predicting chemical trends in data and predicting qualitative phenomena—such as suggesting the conditions under which a deep resonance should descend into the band gap and become a bound deep level.

Because of the intrinsic limitations of contemporary theory, HJALMARSON *et al.* constructed a theory of deep levels that considered only the central-cell impurity potential of the defect. The theory can be and has been modified to include lattice relaxation, charge state splittings and the long-ranged part of the screened Coulomb interaction, but the requirement of simplicity is best met with the Hjalmarson model. In fact, the model's predictions have turned out to be in remarkably good agreement with the data.

In 1969 LANNOO and LENGLART [18] predicted the deep energy level of the diamond vacancy using a simple tight-binding model. Their approach predicts deep levels in good agreement with the most recent calculations [19-26], and their tight-binding ideas provided an essential guide for the development of the Hjalmarson theory. The two elements that are missing from that early work are i) an accurate treatment of both the sp^3 character of the chemical bond and the indirect conduction band structure and ii) a quantitative prescription for predicting the deep levels of impurities as well as vacancies, that is, a scheme for determining the defect potential of an impurity. (For a vacancy the defect potential is infinitely positive, as shown by LANNOO and LENGLART, causing the defect « atom » to be decoupled from the host by virtue of the

infinite-energy denominators in perturbation theory.) Subsequent tight-binding theories of deep impurity levels met with varying degrees of success, but a significant improvement occurred as a result of the work of Vogl *et al.*, which produced the sp^3s^* Hamiltonian with manifest *chemical trends* in its parameters and with adequate conduction bands.

3·3. *Hjalmarson theory of deep levels.* – The Hjalmarson theory of deep impurity levels is a Green's function theory of the type proposed originally by KOSTER and SLATER [27]. The host Hamiltonian matrix H^0 is the Vogl Hamiltonian, eq. (4). Because lattice relaxation and changes of bond length have been neglected, the defect potential matrix V, in the sp^3s^* basis localized at each site, is zero except at the impurity site and is diagonal at the impurity site:

$$(nbR|V|n'b'R') = \delta_{b,b'}\delta_{R,R'}\delta_{n,n'}v_n. \tag{8}$$

Here δ is the Kronecker delta-function and we have $\boldsymbol{v} = (V_s, V_p, V_p, V_p, 0)$; V_s and V_p are given by eqs. (6). The Hamiltonian is

$$H = H^0 + V. \tag{9}$$

Because the defect potential matrix is localized, a Green's function method is useful. Normally to find the impurity levels in a crystal of N unit cells with two atoms per cell and five orbitals per atom, one must solve by brute force a $10N \times 10N$ matrix equation. With a Green's function method, one need only solve a 4×4 matrix: the size of the defect rather than the size of the crystal. In fact, tetrahedral point-group symmetry reduces this 4×4 matrix to four 1×1 matrix equations.

The Green's function matrix for the perturbed crystal is

$$G(E) = (E - H)^{-1} \tag{10}$$

and is related to the unperturbed Green's function

$$G^0(E) = (E - H^0)^{-1} = \sum_{k\lambda} |k\lambda\rangle (E - E_{k\lambda})^{-1} \langle k\lambda| \tag{11}$$

by Dyson's equation

$$G = G^0 - G^0 V G. \tag{12}$$

(Verify this by multiplying on the left by $E - H^0$ and on the right by $E - H$.) The formal solution of Dyson's equation is the matrix

$$G = [1 - G^0 V]^{-1} G^0, \tag{13}$$

which has nontrivial solutions at the energies determined by the secular equation

$$\det |1 - G^0 V| = 0 . \tag{14}$$

Taking matrix elements of the secular equation in the $|nbR)$ basis, we find the eigenvalue equations for the deep-level energy E

$$V_s^{-1} = (sD|G^0|sD) = P \sum_{k\lambda} |(sD|k\lambda)|^2/[E - E_{k\lambda}] \tag{15}$$

and

$$V_p^{-1} = (p_\tau D|G^0|p_\tau D) . \tag{16}$$

Rewriting these results in terms of the local host spectral densities $D^0_{A_1}(E')$ and $D^0_{T_2}(E')$, we have [28]

$$V_s^{-1} = P\int_{-\infty}^{\infty} dE' D^0_{A_1}(E')/[E - E'] \tag{17}$$

and

$$V_p^{-1} = P\int_{-\infty}^{\infty} dE' D^0_{T_2}(E')/[E - E'] \tag{18}$$

with

$$D^0_n(E') = (nD|\delta(E' - H^0)|nD) = \sum_{k\lambda} |(nD|k\lambda)|^2 \delta(E' - E_{k\lambda}) . \tag{19}$$

To see how these results, eqs. (17) and (18), are obtained more directly, it is useful to take matrix elements of Dyson's equation (12) in the $|n\lambda R)$ basis and to consider the impurity site $bR = D$; we have

$$(nD|G|n'D) = (nD|G^0|n'D) + \sum_r (nD|G^0|rD)\, c_r (rD|G|n'D) , \tag{20}$$

where n, n' and r range over s, p_x, p_y and p_z. Because H^0 is invariant under the operations of the tetrahedral point group T_d, H^0 and $(E - H^0)^{-1}$ are invariant operators. Since the s-state transforms according to the A_1 irreducible representation of T_d and the p-states transform according to the x, y and z rows of the T_2 representation, $(nD|G^0|n'D)$ is diagonal in n and eqs. (17) and (18) follow.

The energy E is always to be interpreted as having an infinitesimal positive imaginary part $i0$; this gives the correct boundary conditions for the Green's function. Because of this and the identity $(x - i0)^{-1} = P(1/x) - i\pi\delta(x)$, we have [28]

$$G(E) = P(E - H)^{-1} - i\pi\delta(E - H) ,$$

where P denotes a principal value and $\delta(x)$ is Dirac's delta-function.

Equations (17) and (18) are the central equations of the Hjalmarson theory. To solve them, one needs to first evaluate the spectral densities $D_n^0(E')$ for the host. When the Hamiltonian H^0 is diagonalized to find the eigenvalues $E_{k\lambda}$, the overlap integrals $\langle nbk|k\lambda\rangle$ are the components of the normalized eigenvectors. In terms of these quantities, we have

$$\langle nb\boldsymbol{R}|k\lambda\rangle = N^{-\frac{1}{2}} \exp\left[i\boldsymbol{k}\cdot(\boldsymbol{R}+\boldsymbol{v}_b)\right]\langle nbk|k\lambda\rangle\,.$$

Hence we have

$$D_n^0(E') = N^{-1}\sum_{k\lambda}|\langle nbk|k\lambda\rangle|^2\delta(E'-E_{k\lambda})\,,$$

which can be summed either using the Lehmann-Taut method [29] or (for the deep-level energy E in the fundamental band gap) using the special-point method [30].

In practice, the deep levels $E_{A_1}(V_s)$ and $E_{T_2}(V_p)$ are calculated by computing the functions $V_s(E_{A_1})$ and $V_p(E_{T_2})$ as follows: i) a value of E is selected, ii) the spectral densities $D_n^0(E')$ are evaluated, iii) the right-hand sides of eqs. (17) and (18) are evaluated, and iv) V_s and V_p are determined from those equations. Plots of E *vs.* V for A_1 and T_2 states then give predictions of deep levels *vs.* defect atomic energy (see eqs. (6)). The vacancy levels are the asymptotes $E(V\to\infty)$ of these curves.

Problem 4. Compute $V(E)$ for energies E outside the host band, in the case of a defect in a one-dimensional nearest-neighbor tight-binding crystal. (Hint: $H^0 = \beta\sum_n[|R\rangle(R+1| + |R+1\rangle(R|]$, $V = V_0|D\rangle(D|$. Compute the band structure E_k. Then compute $(R|G^0|R') = N^{-1}\sum_k (E-E_k)^{-1}\exp[ik[R-R']]$. To evaluate $(R|G^0|R') = (R-R'|G^0|0)$ analytically for energies outside of the band, use a contour integral over the unit circle.) Repeat this calculation for the defects on the P site in GaP, using first one special point [30] and then ten special points to evaluate the sums over k. You will obtain good results with ten special points.

4. – Qualitative physics.

The *qualitative* physics determining deep levels is depicted for the case of GaP:N_P in fig. 4, after [11]. In this figure we consider, for simplicity, only the s-states of the atoms (and the A_1-symmetric defect level) and note that the Ga atomic energy ε_{Ga} lies above the P energy ε_P. When these two widely separated atoms are brought together into a molecule, the levels repel—resulting in a bonding-antibonding splitting that, in lowest order of perturbation theory about the infinite-lattice-constant limit, is proportional to $v^2/(\varepsilon_{Ga}-\varepsilon_P)$, where

r is the Ga-to-P transfer matrix element (and is about the same for all semiconductors [7]). The important point is that the bonding-antibonding splitting is inversely proportional to the energy denominator $\varepsilon_{Ga} - \varepsilon_P$. When these molecules are brought together into a solid, the antibonding states produce the conduction band and the bonding states yield the valence band, with the fundamental band gap in between.

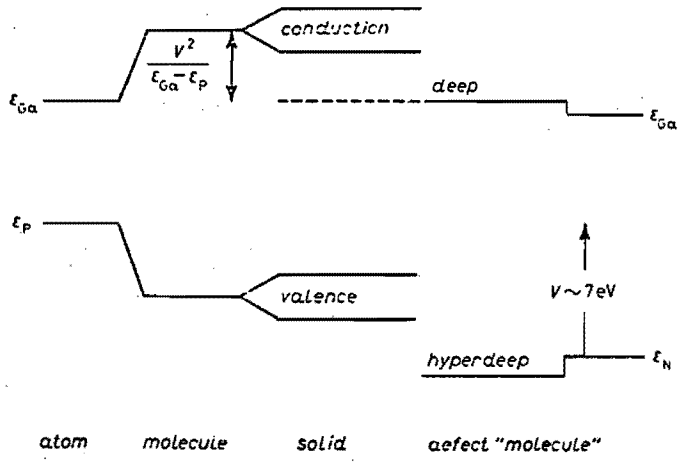


Fig. 4. – Schematic illustration of the qualitative physics governing deep levels, after ref. [3, 11]. See text.

Now imagine a «defect molecule» with a N impurity replacing one of the P atoms. Its atomic energy is $\simeq 7$ eV lower than that of P, and so, when it interacts with a Ga atom to form a molecule, the resulting bonding-antibonding splitting is smaller than for Ga and P, because the energy denominator $\varepsilon_{Ga} - \varepsilon_N$ is $\simeq 7$ eV larger than the denominator $\varepsilon_{Ga} - \varepsilon_P$. As a result, the level in the gap (the deep level) can lie below the conduction band edge in the gap. In fact, fig. 4 illustrates that the issue of whether a deep level is «bound» in the gap or «resonant» with the host bands depends primarily on whether the bands are broad enough to cover up the deep level; that is, it depends on the amount of bonding-antibonding splitting.

Several features of fig. 4 are worthy of special mention: i) The N deep level is derived from the Ga dangling-bond energy ε_{Ga} and is antibonding and *host-like* (Ga-like), not impuritylike (N-like). The N-like level is the bonding *hyperdeep level* lying below the valence band; it is electrically inactive, being full of electrons, and is normally unobserved. ii) The deep level is orthogonal to the hyperdeep level. iii) The deep level is repelled upward by the hyperdeep level by means of the bonding-antibonding level repulsion. iv) The deep level is «pinned» to the Ga dangling-bond level and cannot be pulled below it: Imagine decreasing the energy of the N level ε_N relative to ε_P from $\simeq -7$ eV to $\simeq -15$ eV

(oxygen) and then to $\simeq -1000$ eV (an ideal vacancy): the deep-level energy will move down only slightly, never becoming deeper than the Ga dangling bond or ideal P vacancy energy of ε_{Ga}. This is the meaning of «deep-level pinning»: *Major changes in the deep-level potential result in only minor changes of the deep-level energy*, or $|dE/dV| \ll 1$.

The deep-level pinning can be illustrated by plotting the deep level in the gap *vs.* the defect potential V, which (according to the Vogl model's scaling rules) is proportional to the difference in atomic energies of the defect and the host (P atom). This is done in fig. 5 for the *s*-like A_1 states of defects substituting for P in GaP. The curve $E(V)$ is similar to a hyperbola, having the energy of a Ga dangling bond or a P vacancy as its asymptote. One can see that $E(V=\infty)$ corresponds to a vacancy, because, as the magnitude of the defect potential increases, the defect atom becomes less and less coupled to the host (recall that in perturbation theory the coupling is inversely proportional to an energy denominator of order V) until for $V=\infty$ the defect is totally uncoupled, namely a vacancy [18]. Once one recognizes that the physics of deep levels results in a hyperbolalike curve $E(V)$, the problem of predicting deep

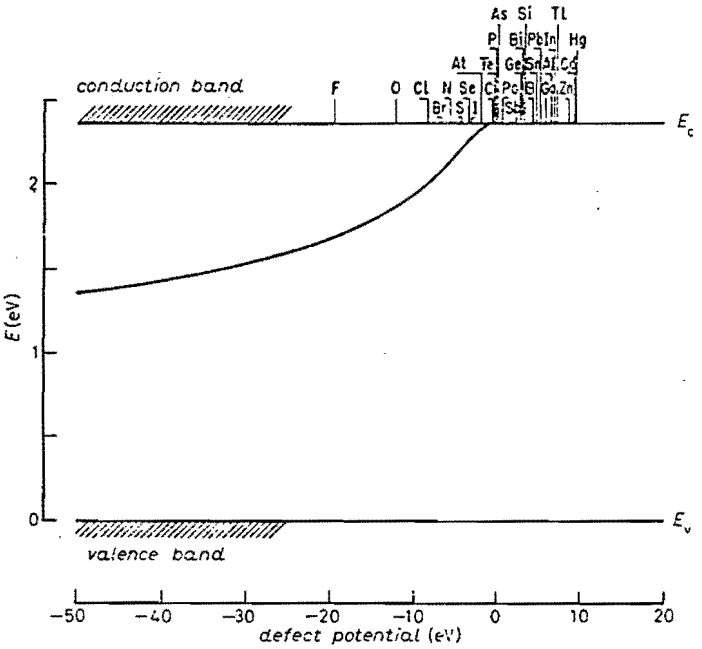


Fig. 5. – Energy levels in the band gap, as calculated by HJALMARSON *et al.* [11], *vs.* defect potential for A_1 symmetric states of defects on the P site in GaP, after ref. [3]. Note that, if the theory is taken literally with no allowances for a theoretical uncertainty, the S defect is predicted to have a deep level in the gap, just slightly below the conduction band edge. Experimentally it is known that S is a shallow donor; hence, one must make allowances for the uncertainty in the theory and recognize that in fact the deep level for S must lie slightly above the conduction band minimum in GaP.

levels reduces to predicting two numbers: the ideal-vacancy energy $E(\infty)$ and the threshold potential V_T at which the resonant deep level passes into the gap and ceases being a resonance.

Figure 6 illustrates very schematically the wave functions of deep levels, using only s-states, for simplicity. The host valence band of GaP has a bonding wave function that is largely P-like, but with a significant Ga-like component.

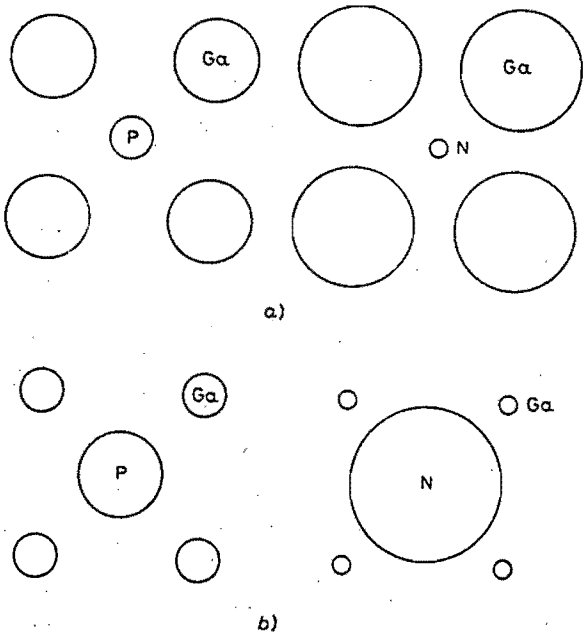


Fig. 6. – Schematic illustration of the wave functions in the (bonding, *a*)) valence and (antibonding, *b*)) conduction bands of GaP (using a s-state model) and in the bonding (hyperdeep) and antibonding (deep) levels of N on a P site in GaP, after ref. [3].

When N replaces P, N is more electronegative and attracts electrons to it. As a result, the bonding hyperdeep level has a wave function that is overwhelmingly N-like, with just a small Ga component. The antibonding deep level is orthogonal to the hyperdeep level, and so has a wave function that is almost exclusively Ga-like, with only a small N component.

All of these ideas have been abstracted from the Hjalmarson *et al.* theory.

5. – Evidence supporting the theory.

The theory of Hjalmarson *et al.* [11, 12] has made literally thousands of predictions of deep levels and can account for an extremely large body of data. Here we review a few representative predictions of the theory.

5·1. *Wave functions.* – Figure 7 shows the magnitude of the deep-level wave function of substitutional S^+ in Si, as a function of distance R from the S center, in comparison with electron nuclear double resonance (ENDOR) data [31, 32]. The ENDOR measurements give the charge density $|\psi(R)|^2$ at Si sites adjacent to and nearby the S defect. The data do not reveal the phase of the wave function, and so we have plotted only the magnitude, even though the wave function itself oscillates rather rapidly. Note that the simple Hjalmarson theory is in excellent agreement with the data, for distances R out to the sixth nearest-neighbor shell. Beyond this distance, the effective-mass wave function (which fails badly for small R) describes the data well—indicating that the present theory would have been in even more dramatic agreement with the data if a Coulomb tail, $-e^2/\varepsilon r$ for $r > a_L$, had been added to the defect potential. This success is by no means trivially obtained, since some theories predict quite inaccurate wave functions for the S^+ deep level [31, 33].

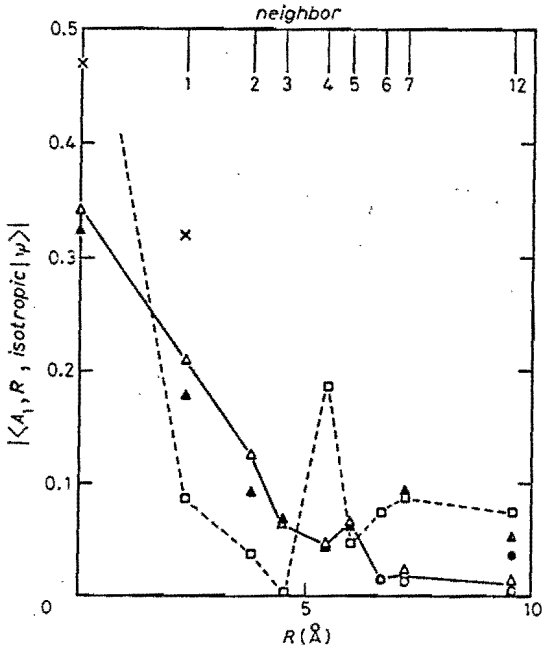


Fig. 7. – The magnitude of the isotropic part of the wave function of a S^+ impurity in Si, as a function of the distance R (in Å) from the impurity site, after ref. [31]. The solid triangles and circles are derived from ENDOR data of ref. [32]; the open triangles and circles are the calculations of ref. [31]. Effective-mass theory is denoted by the dashed line and open squares. The ×'s are the theory discussed in ref. [33].

5·2. *Deep levels in* II-VI *compound semiconductors.* – The sort of predictions the theory produces are illustrated in fig. 8, where the levels associated with column V defects on a S site in CdS are given. These impurities, which one

might naively expect to be shallow acceptors, are, with the exception of N, deep levels predicted to lie well within the band gap of CdS. This is undoubtedly one reason (but not the only reason) why CdS and many other II-VI materials cannot be doped *p*-type—the expected shallow acceptors are, in fact, deep. The exception to this rule is N, which is predicted to yield no deep levels in the gap and to be a shallow acceptor level. Interestingly enough, in ZnSe (another II-VI host), WU *et al.* [34] have ion-implanted N, to find that it does produce a shallow acceptor level—as predicted by KOBAYASHI *et al.* [35].

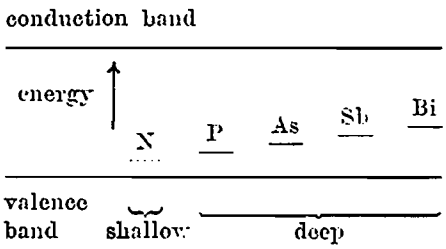


Fig. 8. – Theoretical predictions of ref. [35] showing that the expected standard *p*-type shallow dopants, except N, on the S site in CdS produce deep levels in the gap, after ref. [3].

5·3. *Sankey's theory of paired defects.* – SANKEY *et al.* [36] have extended the theory of deep levels associated with substitutional point defects to pairs of sp^3-bonded substitutional defects—with physically transparent results. The paired defect is a «molecule» that has «molecular orbitals» corresponding to π-like (or e-symmetric) states and σ-like (or a_1-symmetric) states (see fig. 9). The relevant (four) π-like molecular orbitals are composed of T_2 single-defect orbitals polarized perpendicular to the spine of the molecule and have the same energies as the single-defect *p*-like T_2 states. The two *p*-like T_2 states of the single defect that are polarized parallel to the spine of the molecule hybridize

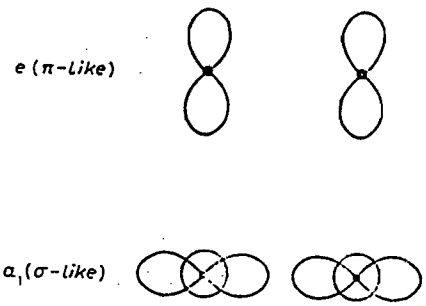


Fig. 9. – Schematic illustration of the π-like e-symmetric molecular orbitals and the σ-like a_1 states of a defect pair. The dots indicate defects, circles denote *s*-states, and propellers denote *p*-states, after ref. [3].

with the two s-like A_1 states centered on the different defects to form σ-like c-symmetric molecular states. In contrast with the π-like states, these hybridizing σ-like states are significantly perturbed by the presence of the second defect. Their energies, however, obey an approximate interlacing theorem, which requires that the «molecular» levels interlace the «atomic» A_1 and T_2 levels. (Recall that the p-like T_2 isolated-defect level decomposes into a σ-like a_1 molecular level polarized along the molecular spine plus two π-like e-levels. Hence both A_1 and T_2 isolated-defect levels produce the same molecular a_1-symmetry σ-like states of the pair. The interlacing theorem states that levels of the same symmetry, namely a_1, when perturbed, do not cross the unperturbed levels.) Because of this interlacing theorem, it is often possible to estimate the energies of the paired-defect levels relative to the isolated-defect levels to within a few tenths of an eV—without executing a calculation.

SANKEY has developed these ideas and applied them to the nearest-neighbor (spectator, oxygen) pairs in GaP (fig. 10). He has shown that the isolated oxygen

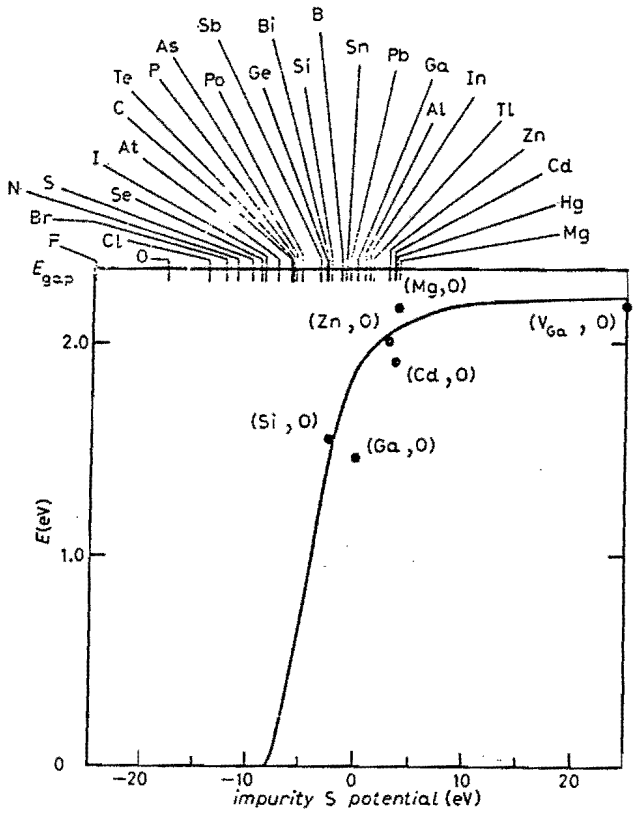


Fig. 10. – Trends in the energy levels for nearest-neighbor paired complexes of a spectator impurity on a Ga site and an O defect on a P site in GaP, after ref. [36]. The dots are data, and the spectator atoms label the abscissa.

level (*i.e.* the (Ga, O) pair) cannot be driven into the conduction band by any electropositive defect on an adjacent cation site, but that it can be driven into the valence band by any one of the following impurities on a neighboring Ga site: F, O, Cl, Br, N, S, Se, or I.

5·4. *Surface defects and Schottky-barrier heights.*

5·4.1. Core excitons at surfaces. Some of the best evidence supporting the theory comes from core exciton experiments, because, by the optical alchemy approximation or $Z+1$ rule [37], a core exciton is identical to an impurity atom [38]: for example, core-excited Ga is Ga plus a core hole plus an electron and (because the core hole has almost the same charge distribution as a proton) is virtually identical to unexcited Ge, the atom immediately to

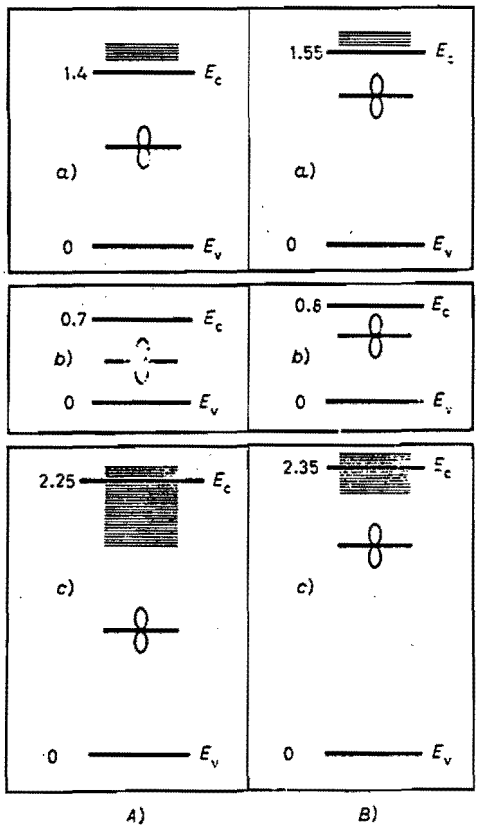


Fig. 11. – Comparison of *A*) experiment and *B*) theory for Ga site (110) surface core excitons in *a*) GaAs, *b*) GaSb and *c*) GaP, after ref. [39]. E_v and E_c denote valence and conduction band edges. The propeller denotes the core exciton level. The horizontal lines denote the lower portion of the intrinsic surface state bands. The theoretical band gaps are appropriate for 4 K, and hence are larger than the experimental gaps obtained at room temperature.

its right in the periodic table. Thus core-excited Ga is a Ge impurity and core-excited In is Sn. Figures 11 and 12 show that the predicted spectra for core-excited Ga at the (110) surface of Ga-group V compounds and for excited In at the surface of In-group V semiconductors account for the data for core

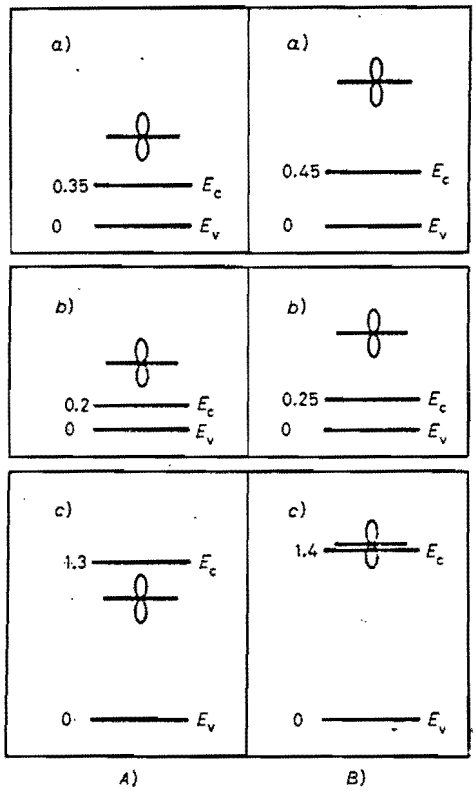


Fig. 12. – Comparison of *A*) experiment and *B*) theory for the In site (110) surface core excitons in *a*) InAs, *b*) InSb and *c*) InP, after ref. [39].

excitons at the relaxed (110) surfaces of these materials [39]. The theory also predicts a transition from shallow effective-mass exciton behavior to deep exciton behavior for the Si $2p$ core exciton in Si_xGe_{1-x} alloys. Evidence of this has been reported very recently by BUNKER *et al.* [38].

5·4.2. Defects at surfaces. An impurity at a surface has energy levels very similar to those of an (impurity, vacancy) pair, because the surface (in a nearest-neighbor tight-binding model) can be created by inserting a sheet of vacancies into the bulk, so that the impurity and the vacancy next to it form a pair whose energy levels are only slightly perturbed by the more distant vacancies of the sheet. To be sure, one must account for lattice relaxation at

the surface; nevertheless, the essential qualitative physics of defects at surfaces is the same as for (defect, vacancy) pairs. Therefore, the deep levels of surface defects can be computed using eq. (14), and the basic ideas used for deep levels associated with substitutional defects carry over to the surface defect problem, where the secular equation is formally the same as eq. (14), but its evaluation is considerably more complicated due to the reduced symmetry of the defect.

One point that should be emphasized is that a surface is a large perturbation, and the *deep levels associated with a surface impurity are likely to lie several tenths of an eV distant from the corresponding bulk impurity levels.* Indeed the number of deep levels bound in the gap may be different for a surface defect from the number for the same defect in the bulk. In particular, impurities that are «shallow» in the bulk often produce one, or even two, deep levels in the gap when they reside at the surface. For example, a nearest-neighbor pair of P impurities in Si is predicted to produce a deep level, although isolated P is a classic shallow donor [36].

5'4.3. Schottky-barrier heights.

5'4.3.1. Bardeen's model of Fermi-level pinning. In 1947 BARDEEN [40] proposed that the Schottky barriers that occur at metal/semiconductor interfaces are due to Fermi-level pinning by states at the interface. Stated simply

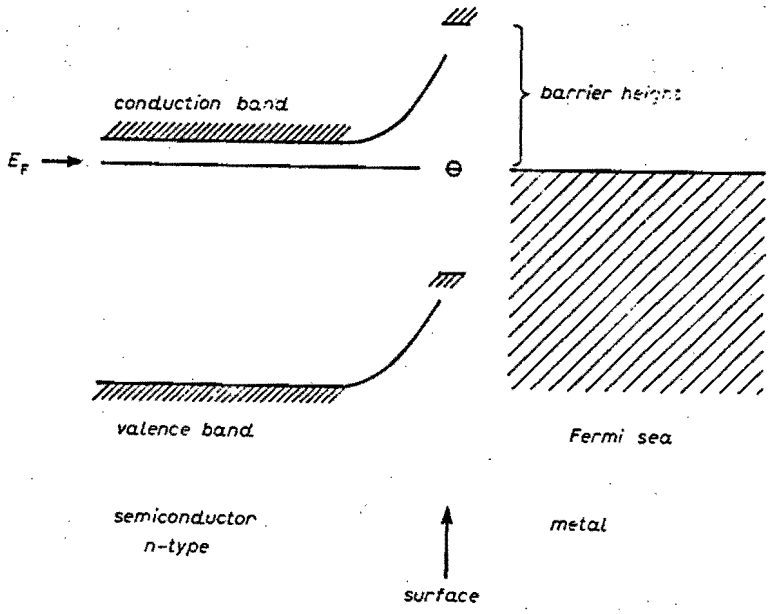


Fig. 13. – Schematic illustration of band bending and Schottky-barrier formation in the Bardeen model, after ref. [3].

for a degenerate *n*-type semiconductor, the Fermi levels of the bulk semiconductor, the metal and the semiconductor surface align in electronic equilibrium. At zero temperature, the Fermi level of the semiconductor surface is the level into which the next electron falls, and, if this is a bound deep level in the gap associated with a surface defect, charge will flow and the host energy bands will bend until this level aligns with the Fermi level of the bulk semiconductor (fig. 13). This causes a Schottky barrier to form.

5'4.3.2. Spicer's native-defect model. SPICER and co-workers [41] have championed the notion that the Bardeen states responsible for Fermi-level pinning of III-V semiconductors are deep levels associated with native surface defects. In this model, the Schottky-barrier height for *n*-type material is the binding energy of the surface deep level with respect to the conduction band edge (see fig. 13).

5'4.3.3. Allen's theory of Schottky-barrier heights in III-V semiconductors. ALLEN *et al.* [42, 43] have calculated the binding energies of deep levels produced by various defects, native antisite defects in particular, at the (110) surfaces of III-V compounds and ternary III-V alloys. This approach followed an earlier suggestion by DAW *et al.* [44] that Fermi-level pinning by deep levels associated with surface vacancies might account for many Schottky-barrier

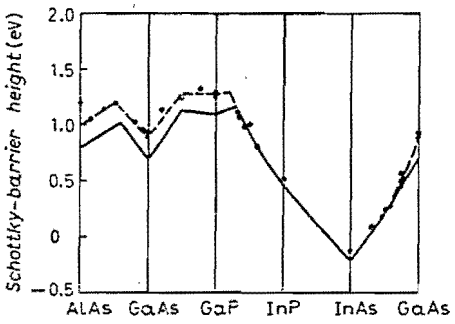


Fig. 14. – Predicted and observed Schottky-barrier heights in III-V alloys *vs.* alloy composition, after ref. [46]: theory (GaAs), ——— experiment (Au). The theory assumes Fermi-level pinning by a cation on an anion site at the Au/semiconductor contact.

height data. The results of Allen's calculations for deep levels associated with surface cation-on-anion-site impurities [45, 46] are given in fig. 14, where they are in remarkable agreement with the data. This simple theory, which has been discussed in detail elsewhere [45], is capable of explaining numerous once-puzzling experimental facts. Fermi-level pinning by antisite defects accounts for the Schottky barriers between nonreactive metals and most III-V

semiconductors. Dangling bonds or extrinsic defects become important when reactive metals or special surface preparations are involved [43]. SANKEY *et al.* [47] have shown recently that Bardeen's ideas can be applied to Si as well as to the III-V's: Schottky-barrier data for Si/transition metal silicide interfaces can be understood in terms of Fermi-level pinning by interfacial dangling bonds. This work unifies the understanding of Schottky-barrier heights of III-V semiconductors and Si and also explains why free-surface calculations give good estimates of the Fermi-level pinning positions of interfacial defects. Hence it appears likely that the Schottky barriers on the common semiconductors have heights determined by surface deep levels (most associated with various native defects).

5·5. *Intrinsic surface states.* – In a nearest-neighbor tight-binding model, a simple way to create a surface is to insert a sheet of vacancies into a bulk semiconductor. The deep levels associated with this sheet defect are surface states. Therefore, the basic theoretical approach to the deep-level problem applies to predicting surface states as well. There are, of course, many technical problems associated with efficiently solving the secular equation (14) for surface states at relaxed and reconstructed surfaces. Those difficulties are beyond the

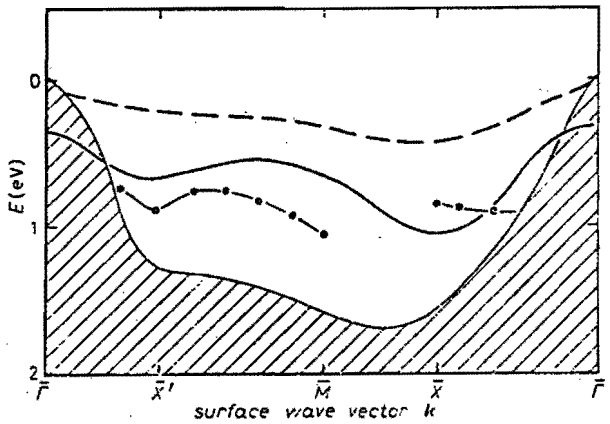


Fig. 15. – Surface state dispersion relations as measured [49] and predicted [48-50] for InP, after ref. [49] and [3]: ——— self-consistent pseudopotential, ------ tight binding, • data.

scope of the present work [45]. Nevertheless, calculations of intrinsic surface state dispersion curves have been executed based on this model [48]. Typical results are given in fig. 15 for the relaxed (110) surface of InP [49]. As with the predictions of deep levels associated with point defects, the surface state predictions are comparably accurate with the best local-density theories [49, 50].

6. – Metastable alloys.

The ideas developed for treating band structures and impurity levels can be applied to treating large concentrations of «impurities» and to predicting phase transitions in alloys. Recently GREENE *et al.* [51] have grown metastable, sustitutional, crystalline $(GaAs)_{1-x}Ge_{2x}$ and $(GaSb)_{1-x}Ge_{2x}$ alloys, even though the constituents, GaAs and Ge or GaSb and Ge, are immiscible at equilibrium. The $(GaAs)_{1-x}Ge_{2x}$ alloys exhibit a V-shaped bowing of the fundamental band gap as a function of alloy composition x (fig. 16). This bowing cannot be

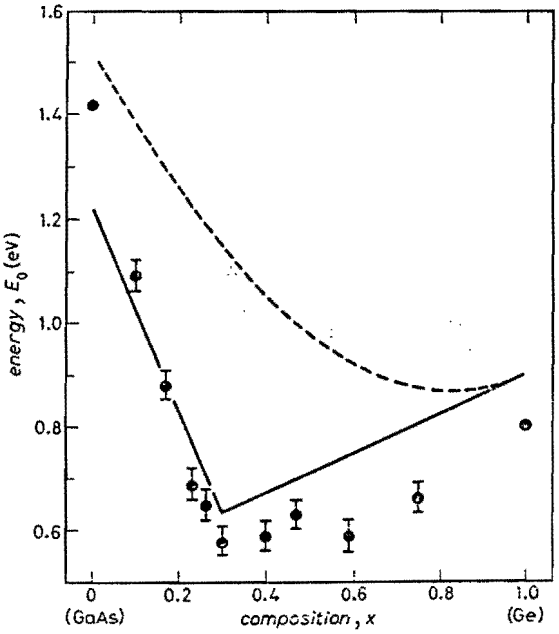


Fig. 16. – Direct energy gap of $(GaAs)_{1-x}Ge_{2x}$ alloys *vs.* alloy composition x, after ref. [53]. The dashed line represents the ordinary virtual-crystal approximation; the solid line represents the theory of ref. [52, 53]. The data are indicated by circles. The theory uses parameters appropriate to 4 K, whereas the data are for room temperature (at which the band gaps are different).

explained by the conventional virtual-crystal approximation, which assumes that each alloy is a crystal whose tight-binding parameters are interpolated between those of GaAs and Ge. NEWMAN *et al.* [52, 53] have shown that the V-shaped bowing is due to an order-disorder transition between a zincblende and a diamond phase (see fig. 17).

In modeling the phase transition in these alloys, NEWMAN was faced with data that required a reasonably accurate theory of electronic structure valid

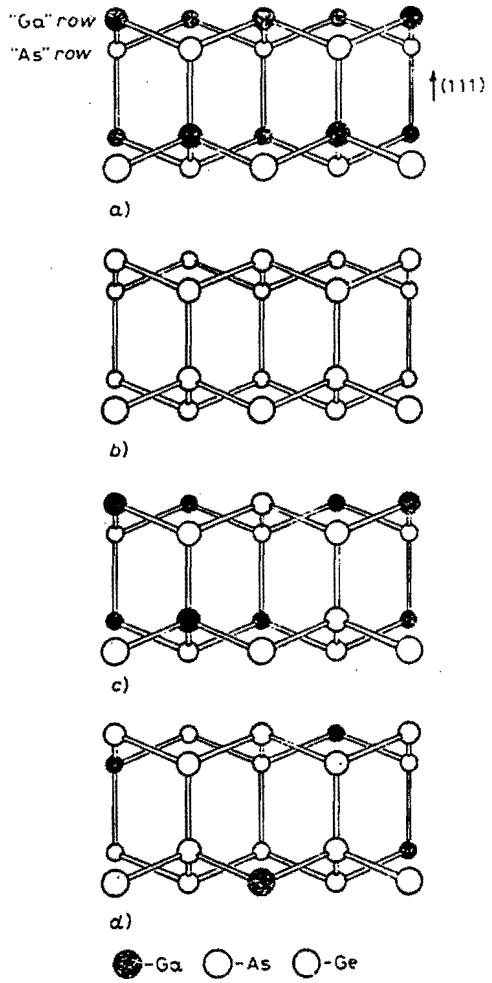


Fig. 17. – A schematic model of the phases of $(GaAs)_{1-x}Ge_{2x}$, after ref. [52, 53]: *a*) the GaAs (ordered) zincblende phase, *b*) the Ge (disordered) diamond phase, *c*) the GaAs-rich ordered zincblende phase of the alloy and *d*) the Ge-rich disordered diamond phase of the alloy. In the ordered phase, the great majority of Ga atoms occupy nominal Ga sites, but in the disordered phase Ga atoms show no site preference.

in both the ordered zincblende and the disordered diamond phases. Yet few theories are capable of accurately treating both a phase transition and electronic structure. NEWMAN circumvented this problem by adopting a two-Hamiltonian model.

The first Hamiltonian involves a pseudospin formalism, in which occupation of a site $\boldsymbol{R}$ by Ga, As, or Ge is represented by a pseudospin $\hat{S}_R$ that is «up» $(+1)$, «down» (-1), or «zero» (0):

$$H = \sum_{R,R'} [J S_R S_{R'} - K S_R^z S_{R'}^z] + \sum_R [h S_R + \Delta S_R^z] .$$

Here J, K, h and Δ are parameters defined in ref. [52], and $\boldsymbol{R}$ and $\boldsymbol{R}'$ are nearest-neighbor sites. In this pseudospin language, crystalline GaAs is a « pseudoantiferromagnet », because alternating sites are occupied by Ga (« up ») and As (« down »); the Ge atoms (« spin zero ») dilute the « antiferromagnetism ». The « spin-spin » interaction between nearest neighbors causes As atoms to preferentially surround Ga and discourages the formation of clusters of only Ga or only As. Details of the Hamiltonian, which is similar to the Blume, Emery, Griffiths Hamiltonian [54] for ^{3}He-^{4}He solutions, are given in ref. [52]. NEWMAN solved this Hamiltonian, in a mean-field approximation, to obtain the equation for the order parameter $M(x, x_c)$:

$$M/(1-x) = \mathrm{tgh}\,[M/(1-x_c)]\,.$$

Here x_c is the alloy composition at which the minimum band gap is observed, 0.3 for $(GaAs)_{1-x}Ge_{2x}$. The order parameter is related to the probability that a Ga atom will be found on a nominal cation site of an imagined zincblende lattice; if this probability is the same as that for finding Ga on a nominal anion site, then the crystal structure is « diamond » and not « zincblende ». That is, the alloy is « disordered ».

The second Hamiltonian is the Vogl empirical tight-binding theory, but with the alloy's parameters determined by a new virtual-crystal approximation that depends on the order parameter of the first Hamiltonian: The diagonal matrix elements are interpolated assuming, for example, that the average cation is $[(1-x+M)/2]\,\mathrm{Ga} + [(1-x-M)/2]\,\mathrm{As} + [x]\,\mathrm{Ge}$. The energy band structure computed using this model yields a direct band gap $E_0(x)$ *vs.* alloy composition x (fig. 16) and explains the observed V-shaped bowing.

7. – Summary.

In summary, the simple ideas originating in the work of Hsu, Wolford and Streetman and quantified in both the Vogl empirical tight-binding scheme and the Hjalmarson theory of defects have proven to have widespread applicability to defect states, interface states, surface states, Schottky barriers and alloy theory. The most redeeming feature of the theory is that it is simple enough to be used by a nonexpert, and yet it produces rather good predictions for the « deep » electronic states associated with almost any localized perturbation in a semiconductor.

* * *

We are grateful to the U. S. Office of Naval Research and the U. S. Army Research Office for their generous support of this research (Contract Nos. N00014-77-C-0537, N00014-84-K-0357 and DAAG29-83-K-0122). We have

benefitted greatly from many stimulating conversations with colleagues and students throughout the period of this research. The author apologizes to the many authors whose work he has not cited adequately; his sole excuse for this omission is a desire to present lectures from a uniform viewpoint and notation. He would like to thank his co-authors cited below for teaching him so much. In the preparation of this manuscript, he benefitted from conversations with D. FROELICH, D. JENKINS, G. KIM, K. NEWMAN, J. SHEN and D. WOLFORD.

REFERENCES

[1] D. J. WOLFORD and B. G. STREETMAN: *Phys. Rev. Lett.*, **36**, 1400 (1976); D. J. WOLFORD: Ph.D. Thesis, University of Illinois (1979). The actual binding energy of the S impurity in GaP is 107 meV, larger than the 35 meV effective-mass value due to valley-orbit splitting.
[2] D. J. WOLFORD, W. Y. HSU, J. D. DOW and B. G. STREETMAN: *J. Lumin.*, **18/19**, 863 (1979).
[3] J. D. DOW: figures to be published.
[4] The virtual-crystal approximation treats the alloy as a crystal whose anion is the average anion $(1-x)\,\mathrm{As} + x\mathrm{P}$. The off-diagonal matrix elements of a tight-binding model are obtained by linearly interpolating Vd^2, where d is the bond length obtained by linear interpolation.
[5] W. Y. HSU, J. D. DOW, D. J. WOLFORD and B. G. STREETMAN: *Phys. Rev. B*, **16**, 1597 (1977).
[6] P. VOGL, H. P. HJALMARSON and J. D. DOW: *J. Phys. Chem. Solids*, **44**, 365 (1983). Table I of this reference should be corrected as follows: For ZnTe, we have $E(s, c) = -0.9350$ and $V(pa, s^*a) = 1.2381$. For SiC we have $V(pa, sc) = 4.2218$ $(= V(sc, pa))$ and $V(pa, s^*c) = 3.9665$. The 4's in eq. (9) should be 2's.
[7] W. A. HARRISON: *Electronic Structure and the Properties of Solids* (San Francisco, Cal., 1980).
[8] W. A. HARRISON: in *Festkörperprobleme*, Vol. **17**, edited by J. TREUSCH (Wiesbaden, 1977), p. 135.
[9] W. KOHN: *Solid State Phys.*, **5**, 257 (1957).
[10] F. BASSANI and P. PARRAVICINI: *Electron States and Optical Transition in Solids* (Elmsford, 1975).
[11] H. P. HJALMARSON, P. VOGL, D. J. WOLFORD and J. D. DOW: *Phys. Rev. Lett.*, **44**, 810 (1980), and to be published.
[12] H. P. HJALMARSON: Ph. D. Thesis, University of Illinois (1979) (unpublished).
[13] O. F. SANKEY and J. D. DOW: *Phys. Rev. B*, **27**, 7641 (1983).
[14] L. A. HEMSTREET: *Phys. Rev. B*, **15**, 834 (1977); P. VOGL: to be published.
[15] M. SCHEFFLER, J. P. VIGNERON and G. B. BACHELET: *Phys. Rev. Lett.*, **49**, 1765 (1982).
[16] J. BERNHOLC, N. O. LIPARI, S. T. PANTELIDES and M. SCHEFFLER: *Phys. Rev. B*, **26**, 5706 (1982).
[17] S. LEE, J. D. DOW and O. F. SANKEY: *Phys. Rev. B*, to be published.
[18] M. LANNOO and P. LENGLART: *J. Phys. Chem. Solids*, **30**, 2409 (1969).
[19] E. KAUFER, P. PECHEUR and M. GERL: *J. Phys. C*, **9**, 2319 (1976).

[20] J. D. JOANNOPOULOS and E. J. MELE: *Solid State Commun.*, **20**, 729 (1976).
[21] M. JAROS and S. BRAND: *Phys. Rev. B*, **14**, 4494 (1976).
[22] U. LINDEFELT: *J. Phys. C*, **12**, L419 (1979); D. A. PAPACONSTANTOPOULOS and E. N. ECONOMOU: *Phys. Rev. B*, **22**, 2903 (1980).
[23] G. A. BARAFF and M. SCHLÜTER: *Phys. Rev. Lett.*, **41**, 892 (1978); *Phys. Rev. B*, **19**, 4965 (1979).
[24] J. BERNHOLC and S. T. PANTELIDES: *Phys. Rev. B*, **18**, 1780 (1978); J. BERNHOLC, N. O. LIPARI and S. T. PANTELIDES: *Phys. Rev. Lett.*, **41**, 895 (1978); *Phys. Rev. B*, **21**, 3545 (1980); J. BERNHOLC, S. T. PANTELIDES, N. O. LIPARI and A. BALDERESCHI: *Solid State Commun.*, **37**, 705 (1981).
[25] C. A. SWARTS, D. L. MILLER, D. R. FRANCESCHETTI, H. P. HJALMARSON, P. VOGL, J. D. DOW, D. J. WOLFORD and B. G. STREETMAN: *Phys. Rev. B*, **21**, 1708 (1980).
[26] W. A. HARRISON: *Phys. Rev. B*, **8**, 4487 (1973); D. J. CHADI and M. L. COHEN: *Phys. Status Solidi B*, **68**, 405 (1975); K. C. PANDEY and J. C. PHILLIPS: *Phys. Rev. B*, **13**, 750 (1976); D. J. CHADI: *Phys. Rev. B*, **16**, 790 (1977).
[27] G. F. KOSTER and J. C. SLATER: *Phys. Rev.*, **96**, 969 (1951).
[28] For example, the density of states of the perturbed Hamiltonian is

$$D(E') = -\pi^{-1}\,\mathrm{Im}\,\mathrm{Tr}\,G(E')$$

and the density of states of the host is

$$D^0(E') = -\pi^{-1}\,\mathrm{Im}\,\mathrm{Tr}\,G^0(E') = -\pi\sum_{k\lambda}\delta(E'-E_{k\lambda})\,.$$

[29] G. LEHMANN and M. TAUT: *Phys. Status Solidi B*, **54**, 469 (1972); J. RATH and A. J. FREEMAN: *Phys. Rev. B*, **11**, 2109 (1975).
[30] D. J. CHADI and M. L. COHEN: *Phys. Rev. B*, **8**, 5747 (1973).
[31] S. Y. REN, W. M. HU, O. F. SANKEY and J. D. DOW: *Phys. Rev. B*, **26**, 951 (1982). See also S. Y. REN: *Sci. Sin.*, **27**, 443 (1984).
[32] G. W. LUDWIG: *Phys. Rev.*, **137**, A 1520 (1965).
[33] T. SHIMIZU and K. MINAMI: *Phys. Status Solidi B*, **48**, K181 (1971).
[34] Z. L. WU, J. L. MERZ, C. J. WERKHOVEN, B. J. PATRICK and R. N. BHARGAVA: *Appl. Phys. Lett.*, **40**, 345 (1982).
[35] A. KOBAYASHI, O. F. SANKEY and J. D. DOW: *Phys. Rev. B*, **28**, 946 (1983).
[36] O. F. SANKEY, H. P. HJALMARSON, J. D. DOW, D. J. WOLFORD and B. G. STREETMAN: *Phys. Rev. Lett.*, **45**, 1656 (1980); O. F. SANKEY and J. D. DOW: *Appl. Phys. Lett.*, **38**, 685 (1981); *J. Appl. Phys.*, **52**, 5139 (1981); *Phys. Rev. B*, **26**, 3243 (1982); *Solid State Commun.*, in press. See also J. D. DOW and O. F. SANKEY: *Am. Inst. Phys. Conf. Proc.*, Vol. **73**, edited by R. A. STREET, D. K. BEIGELSEN and J. C. KNIGHTS (New York, N. Y. 1981), p. 141.
[37] See, for example, J. D. DOW, D. R. FRANCESCHETTI, P. C. GIBBONS and S. E. SCHNATTERLY: *J. Phys. F*, **5**, L211 (1975).
[38] H. P. HJALMARSON, H. BÜTTNER and J. D. DOW: *Phys. Rev. B*, **24**, 6010 (1981). Experimental evidence supporting this theory has recently been reported by B. A. BUNKER, S. L. HULBERT, J. P. STOTT and F. C. BROWN: *Phys. Rev. Lett.*, **53**, 2157 (1984).
[39] R. E. ALLEN and J. D. DOW: *Phys. Rev. B*, **24**, 911 (1981).
[40] J. BARDEEN: *Phys. Rev.*, **71**, 717 (1947).
[41] W. E. SPICER, P. W. CHYE, P. R. SKEATH, C. Y. SU and I. LINDAU: *J. Vac. Sci. Technol.*, **16**, 1422 (1979), and references therein; W. E. SPICER, I. LINDAU, P. R. SKEATH and C. Y. SU: *J. Vac. Sci. Technol.*, **17**, 1019 (1980), and references therein.

[42] R. E. ALLEN and J. D. DOW: *Phys. Rev. B*, **25**, 1423 (1982).
[43] J. D. DOW and R. E. ALLEN: *J. Vac. Sci. Technol.*, **20**, 659 (1982). See also ref. [45] and references therein.
[44] M. S. DAW and D. L. SMITH: *Phys. Rev. B*, **20**, 5150 (1979); *J. Vac. Sci. Technol.*, **17**, 1028 (1980); *Appl. Phys. Lett.*, **36**, 690 (1980); *Solid State Commun.*, **37**, 205 (1981); M. S. DAW, D. L. SMITH, C. A. SWARTS and T. C. MCGILL: *J. Vac. Sci. Technol.*, **19**, 508 (1981).
[45] For a review of the theory of intrinsic and extrinsic surface states, see J. D. DOW, R. E. ALLEN and O. F. SANKEY: *Chemistry and Physics on Solid Surfaces*, Vol. **5**, edited by R. VANSELOW and R. HOWE, Springer Series in Chemical Physics, **35** (Berlin, Heidelberg, New York, N.Y., 1984).
[46] R. E. ALLEN, T. J. HUMPHREYS, J. D. DOW and O. F. SANKEY: *J. Vac. Sci. Technol. B*, **2**, 449 (1984).
[47] O. F. SANKEY, R. E. ALLEN and J. D. DOW: *Solid State Commun.*, **49**, 1 (1984).
[48] R. P. BERES, R. E. ALLEN and J. D. DOW: *Phys. Rev. B*, **26**, 5702 (1982).
[49] G. P. SRIVASTAVA, I. SINGH, V. MONTGOMERY and R. H. WILLIAMS: *J. Phys. C*, **16**, 3627 (1983).
[50] F. MANGHI, E. MOLINARI, C. M. BERTONI and C. CALANDRA: *J. Phys. C*, **15**, 1099 (1982).
[51] J. E. GREENE: *J. Vac. Sci. Technol. B*, **1**, 229 (1983), and references therein.
[52] K. E. NEWMAN and J. D. DOW: *Phys. Rev. B*, **27**, 7495 (1983).
[53] K. E. NEWMAN, A. LASTRAS-MARTINEZ, B. KRAMER, S. A. BARNETT, M. A. RAY, J. D. DOW, J. E. GREENE and P. M. RACCAH: *Phys. Rev. Lett.*, **50**, 1466 (1983).
[54] M. BLUME, V. J. EMERY and R. B. GRIFFITHS: *Phys. Rev. A*, **4**, 1071 (1971).

Reprinted From
Highlights of
Condensed-Matter Theory

Soc. Italiana di Fisica - Bologna - Italy

Predicted energy band gaps of $(A^{III}B^{V})_{1-x}X^{IV}_{2x}$ metastable, substitutional, crystalline alloys

David W. Jenkins, Kathie E. Newman, and John D. Dow
Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556
(Received 21 January 1985)

Predictions of the energy band gaps as functions of alloy composition are given for the Greene alloys, which are metastable, crystalline, substitutional alloys of III-V compounds and group-IV elemental materials. All possible combinations of these alloys involving Al, Ga, In, P, As, Sb, Si, Ge, and Sn are considered. The Γ and L conduction-band minima, relative to the valence-band maxima, exhibit characteristic V-shaped bowing and kinks as functions of composition x; the band edges at point X bifurcate at critical compositions corresponding to the order-disorder transition of Newman *et al.* The V-shaped bowing due to the transition offers the possibility of band gaps significantly smaller than expected on the basis of the conventional virtual-crystal approximation. Alloys with modest lattice mismatches that are predicted to have especially interesting band gaps include $(InP)_{1-x}Ge_{2x}$, $(AlSb)_{1-x}Sn_{2x}$, $(GaSb)_{1-x}Sn_{2x}$, and $(InAs)_{1-x}Sn_{2x}$, which are alloys with potentially small band gaps, and $(AlAs)_{1-x}Ge_{2x}$ and $(GaAs)_{1-x}Si_{2x}$, which are alloys with larger gaps and several interesting band-edge crossings as functions of composition.

I. INTRODUCTION

Recently, Greene and co-workers have fabricated a new class of semiconducting $(A^{III}B^{V})_{1-x}X^{IV}_{2x}$ alloys for a wide range of compositions.[1–4] The III-V compounds and group-IV elemental materials are normally immiscible at equilibrium,[5] but can be forced to mix by ion bombardment during growth. The resulting material, in the case of $(GaAs)_{1-x}Ge_{2x}$ or $(GaSb)_{1-x}Ge_{2x}$, is a metastable, crystalline, substitutional alloy with a lifetime at room temperature of order 10^{29} years.[6] The fundamental energy band gap of $(GaAs)_{1-x}Ge_{2x}$ has been determined from optical-absorption measurements and shows a nonparabolic V-shaped bowing as a function of alloy composition x (Ref. 7). A V-shaped band gap cannot be explained using the conventional virtual-crystal approximation, which gives approximately parabolic bowing. This V-shaped bowing is explained, however, with a zinc-blende-to-diamond, order-disorder phase transition.[8]

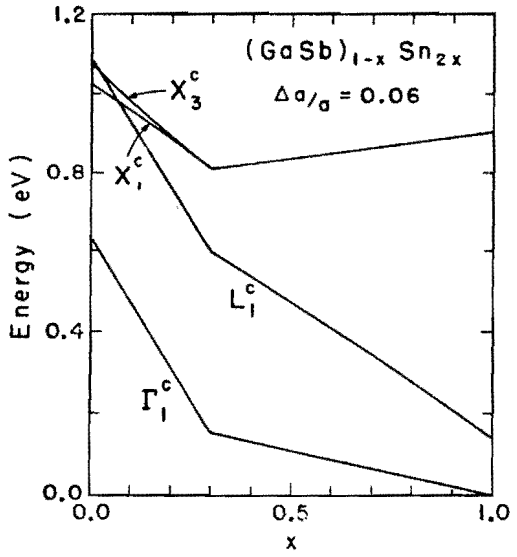


FIG. 1. Predicted band gaps at points Γ, L, and X versus alloy composition for $(GaSb)_{1-x}Sn_{2x}$. Kinks are seen in the Γ and L levels and the level at point X bifurcates at the assumed critical composition of Newman's zinc-blende-to-diamond phase transition, $x_c=0.3$. The gap is direct for all compositions, ranges from $\sim$0.6 to zero and decreases slowly as a function of composition from 0.15 eV to zero for compositions greater than the critical composition.

A theory for this transition has been developed by Newman *et al.*[8–10] and applied to $(GaAs)_{1-x}Ge_{2x}$. As seen in Fig. 1, where the theory is evaluated for the conduction-band minima near points Γ, L, and X for $(GaSb)_{1-x}Sn_{2x}$, the fundamental band gap exhibits a V-shaped bowing as a function of composition, with a kink at the critical composition x_c. This theory also gives smaller gaps than those of the conventional virtual-crystal approximation.

In this paper we apply this theory to the entire class of $(A^{III}B^{V})_{1-x}X^{IV}_{2x}$ alloys involving all possible combinations of Al, Ga, In, P, As, Sb, Si, Ge, and Sn, and we predict the energy band edges for these new metastable materials as functions of alloy composition x. We also establish general rules for understanding the chemical trends in the band gaps and for choosing a metastable $(A^{III}B^{V})_{1-x}X^{IV}_{2x}$ alloy with a desired energy band gap.

II. THEORY

The central idea of the present work is that all of the $(A^{III}B^{V})_{1-x}X^{IV}_{2x}$ metastable alloys should exhibit an order-disorder transition from an ordered zinc-blende structure (in which cations "know" which sites are supposed to be cation sites) to the disordered diamond structure in which there is no distinction between anion and cation sites. The critical composition x_c at which this tran-

sition occurs depends on the growth conditions of the alloy.

In developing a theory of the electronic structures of these alloys, we must remember that very little is presently known about these new and interesting materials. Many of the metastable alloys have not yet been grown; in most cases, satisfactory growth conditions are not yet known; and it is not yet definitely known if any of the Greene alloys other than $(GaSb)_{1-x}Ge_{2x}$ exhibits the order-disorder transition [which should be detected in x-ray diffraction as the disappearance of the (200) zinc-blende spot as x approaches x_c from below].[11] These facts are important in defining the nature of the theory that is appropriate at this time; it should be global and simple, rather than detailed and excessively quantitative. With this in mind, we assume both that all of the Greene alloys exhibit the Newman *et al.* transition, and that there exist growth conditions that will result in a critical composition $x_c=0.3$, the value appropriate for the two alloys grown to date by Greene and co-workers: $(GaAs)_{1-x}Ge_{2x}$ and $(GaSb)_{1-x}Ge_{2x}$ (x_c is probably experimentally adjustable).[12] We then predict the band structures (as functions of alloy composition x) of the remaining $(A^{III}B^{V})X^{IV}_{2x}$ metastable alloys with the intent of determining which alloys are likely to exhibit interesting and useful electronic structures—thereby targeting specific alloys for priority growth. Thus, we present these calculations in order to predict which materials are most likely to be interesting, rather than pretending to specify the band structures with any precision.

A. Order-disorder transition

The order-disorder transition involves a change of symmetry from the zinc-blende structure to the diamond-crystal structure. In this transition, the distinction between anion and cation sites is lost. The relevant order parameter is:[9]

$$M(x)=\langle P_{III}\rangle_{cation}-\langle P_{III}\rangle_{anion}, \tag{1}$$

where we imagine a zinc-blende lattice with sites labeled nominally "cation" and "anion," and $\langle P_{III}\rangle_{cation}$ is the average over all the lattice sites of the probability that a column-III atom occupies a nominal cation site. Thus $M(x)$ is proportional to the average electric dipole moment per unit cell. The order parameter depends on the growth conditions (e.g., substrate temperature, ion-bombardment energy) as well as on the composition x. For a completely ordered zinc-blende alloy, in which all column-III (column-V) atoms occupy nominal cation (anion) sites, we have $M=1-x$. If all the cations are on anion sites and the anions are on cation sites, we have merely mislabeled the nominal lattice and the order parameter is $x-1$. For the metastable ordered phase ($x<x_c=0.3$), we have $0<|M(x)|<1-x$. For the disordered diamond phase ($x>x_c$), we have $M=0$.

The theoretical problem posed by the Greene alloys is that of predicting the electronic structure of metastable alloys which are described by the order parameter $M(x)$. Thus, we must first execute a nonequilibrium phase-transition theory of $M(x)$ and then calculate the changes of the electronic structure as the alloys (with different composition x) undergo the order-disorder transition. Newman showed that this formidable problem could be solved by breaking it into four connected parts: (i) an equilibrium phase-transition theory of the order parameter $M(x)$, based on a three-component "spin"-Hamiltonian model similar to the Blume, Emery, Griffiths model[13] of He^3-He^4 solutions. [Spin-up, spin-down, or zero at a site in $(GaAs)_{1-x}Ge_{2x}$ signifies occupation of that site by Ga, As, or Ge, respectively.] (ii) Introduction of the nonequilibrium character of the alloys by eliminating those equilibrium phases that cannot be reached due to growth conditions (e.g., phase separation, which occurs at equilibrium, is prevented because characteristic growth times are small in comparison with the time required for the phases to diffuse apart); (iii) mutual elimination of two unknown parameters of the spin-Hamiltonian model, i.e., a spin-coupling constant J and an effective growth temperature T, in favor of one empirical parameter, the critical composition x_c;[14] and (iv) evaluation of the electronic structure using a modified virtual-crystal approximation and a tight-binding model[15] whose matrix elements depend parametrically on the order parameter $M(x;x_c)$. Thus, in the Newman approach there are two Hamiltonians: (i) a spin-Hamiltonian for treating the order-disorder transition and for calculating the order parameter $M(x;x_c)$ and (ii) an empirical tight-binding Hamiltonian—that depends parametrically on $M(x;x_c)$—for calculating the electronic structure.

B. Spin-Hamiltonian model

Newman *et al.* have shown that a III-V compound semiconductor such as GaAs can be modeled in a spin-Hamiltonian language as an "antiferromagnet" where spin-up or spin-down on a site represents occupation by a group-III atom or a group-V atom, respectively. Thus GaAs, with alternating Ga and As atoms, in this language, is an "antiferromagnet." The "magnetization" is proportional to the net electric dipole moment per unit cell, Eq. (1), and for zero-temperature GaAs at equilibrium, equals unity. In metastable $(A^{III}B^{V})_{1-x}X^{IV}_{2x}$ alloys such as $(GaAs)_{1-x}Ge_{2x}$, occupation of a site by a column-IV atom such as Ge is represented by "spin" zero. If the Ge were to occupy both anion and cation sites without disturbing the occupation of these sites by Ga and As, then the order parameter would be $M(x)=1-x$. However, M is not $1-x$ because Ge (spin zero) dilutes the "magnetization" $M(x;x_c)$ of this "antiferromagnet," by removing nonzero "spins" at various sites, until there is insufficient "spin-spin" interaction for an average site to "know" it should have spin-up or spin-down. With a sufficient concentration x of dilutants (that depends on temperature), the "magnetization" vanishes, and the system undergoes a phase transition, from an "antiferromagnetic" zinc-blende state with $M\neq0$ to an "unmagnetized" phase ($M=0$). That is, as Ge dilutes GaAs, an average cation site is no longer fully surrounded by As atoms and no longer feels electronically compelled to be occupied by a Ga atom rather than an As atom. The average electric dipole moment $M(x)$ of the ordered zinc-blende phase decreases and the system undergoes a transition from an ordered zinc-blende phase in which Ga atoms preferentially

occupy nominal cation sites to a disordered ($M=0$) diamond phase in which there is no distinction between anion and cation sites. Newman constructed a spin-Hamiltonian model of this order-disorder transition. The important physical parameter of this Hamiltonian is a nearest-neighbor spin coupling (which is related to energies of interaction of the pairs of atoms V-V, III-III, and III-V). The Hamiltonian, when treated in a mean-field approximation, yields the following equation for the order parameter $M(x;x_c)$:

$$\tanh[(M/(1-x_c)]=[M/(1-x)] , \quad (2)$$

where x_c is the critical composition of the order-disorder transition.

C. Tight-binding Hamiltonian

The electronic structure calculations are based on an empirical, ten-band, second-nearest-neighbor, tight-binding theory, which employs an sp^3s^* basis at each site of the zinc-blende lattice. The on-site and nearest-neighbor matrix elements of this model have been obtained previously by Vogl *et al.*,[15] who fit the known band structures of many III-V compounds and group-IV semiconductors. The Vogl matrix elements are augmented by one or two second-neighbor parameters[16] (see Table I) in order to obtain a better fit to the band structures of these semiconductors at the L point of the Brillouin zone. (The Vogl model was designed to fit the conduction-band structures well near points Γ and X.) The on-site matrix elements for these many semiconductors exhibit manifest chemical trends that depend only on the atomic energies of the atom on the site, to a good approximation. The off-diagonal nearest-neighbor matrix elements are inversely proportional to the square of the bond length d, according to the rule of Harrison *et al.*[17] For our purposes the important physical parameters of the tight-binding Hamiltonian are the on-site energies of the column-III, -IV, and -V atoms, which we shall interpolate using a generalized virtual-crystal approximation.[9] The on-site matrix elements are interpolated according to Eq. (3), as are Vd^2, where V is the off-diagonal matrix elements and d is the bond length of the alloy predicted by Vegard's law:[18] $d(x)=(1-x)d_{\text{III-V}}+xd_{\text{IV}}$.

We expect these $(A^{\text{III}}B^{\text{V}})_{1-x}X^{\text{IV}}_{2x}$ alloys to satisfy adequately the Onodera-Toyozawa[19] criterion for an "amalgamated" electronic spectrum, since the variations in on-site diagonal matrix elements are small in comparison with nearest-neighbor transfer matrix elements.[20] Therefore, we expect them to have relatively well-defined band structures which can be described (in a first approximation) by a mean-field theory of the virtual-crystal type. They cannot be treated with the ordinary virtual-crystal approximation, however, because (in the disordered "diamond" phase, in particular) they contain many antisite atoms (e.g., a column-III atom on a nominal anion site)—and the usual virtual-crystal approximation does not allow for antisite atoms. We circumvent this problem by using the generalized virtual-crystal approximation,[9] which has virtual anions and cations such that the virtual cation is (schematically):

$$[(1-x+M)/2]A^{\text{III}}+[(1-x-M)/2]B^{\text{V}}+xX^{\text{IV}} . \quad (3)$$

Here, A^{III}, X^{IV}, and B^{V} represent the column-III, -IV, and -V atoms, and $M(x;x_c)$ is the order parameter (1) of the order-disorder transition, obtained by solving Eq. (2).

TABLE I. Second-neighbor parameters. Note here that $\epsilon(p_xa,p_ya')=\epsilon(p_xc,p_yc')$ and $\epsilon(sa,p_xa)=\epsilon(p_xc,sc')$. See Ref. 16 for details.

Semiconductor	$\epsilon(sa,p_ya')$	$\epsilon(p_xa,p_ya')$
AlP	1.990	0.000
AlAs	1.830	−0.876
AlSb	0.101	0.000
GaP	0.641	0.000
GaAs	0.464	0.000
GaSb	0.688	0.000
InP	0.368	0.000
InAs	0.187	0.000
InSb	0.107	0.000
Si	0.000	0.146
Ge	0.157	0.000
Sn	0.000	0.056

III. RESULTS

The energies of the band edges (relative to the valence-band maximum, which is defined to be the zero of energy) are given in Fig. 1 for $(GaSb)_{1-x}Sn_{2x}$. Corresponding results for all possible $(A^{\text{III}}B^{\text{V}})_{1-x}X^{\text{IV}}_{2x}$ alloys are given in Figs. 2–4. The Γ conduction-band minimum occurs at $k=(0,0,0)$ in the band structure. The edges labeled Δ and Λ refer to the conduction minima near the (1,0,0) and $(\frac{1}{2},\frac{1}{2},\frac{1}{2})$ points, respectively (i.e., near points X and L).[21] For k at the X point of the Brillouin zone, the conduction-band edge actually bifurcates as a function of alloy composition at the critical composition x_c, producing both an X_1 and an X_3 minimum in the zinc-blende (ordered) phase for $x<x_c$, but only one minimum for $x>x_c$ in the diamond (disordered) phase. This bifurcation is reflected in the dependence of the minima along the Δ line as functions of composition x (see Fig. 4), because these minima lie at wave vectors near point X. The relative minimum at point Γ, when plotted as a function of composition x, exhibits a kink at x_c, as does the band edge at the L point. The minimum in the Λ direction reflects the kinked behavior of the nearby L point.

In addition to the dependences on alloy composition x, there are discernible trends depending on the positions of the atoms in the Periodic Table. To facilitate quantification of these trends, we define an effective average atomic number:

$$\langle Z\rangle = xZ_{\text{IV}}+(1-x)(Z_{\text{III}}+Z_{\text{V}})/2 , \quad (4)$$

where, for example, Z_{III} is the atomic number of the column-III atom. Figure 5 shows that the Γ, Δ, and Λ band edges tend to decrease in energy with increasing

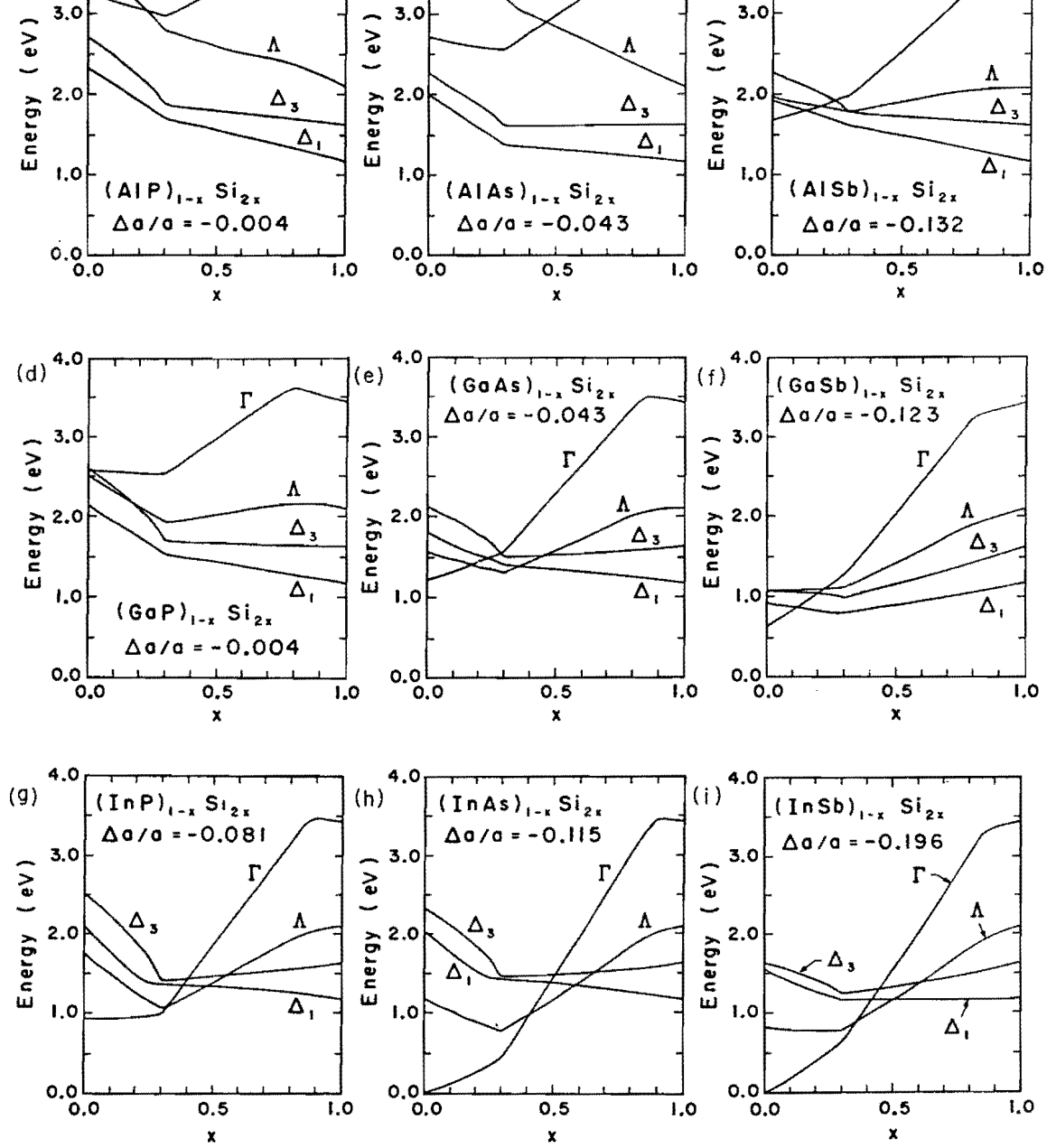


FIG. 2. Predicted band gaps of $(A^{III}B^{V})_{1-x}Si_{2x}$ alloys versus x, for the following III-V compounds: AlP, AlAs, AlSb, GaP, GaAs, GaSb, InP, InAs, and InSb. Lattice mismatches, defined by Eq. (5), are shown. Kinks can be seen in the Γ, Λ, and Δ levels at the assumed critical composition $x_c = 0.3$. The Δ minimum generally lies some distance from the X point in our tight-binding model, so the strict bifurcation at the X point is not clearly visible. The kinks near $x = 1$ are due to a crossing of the Γ_c^{15} and Γ_c^1 levels.

$\langle Z \rangle$, with Γ decreasing most rapidly and Δ decreasing least rapidly with $\langle Z \rangle$. This trend can be exploited, for example, to find metastable alloys with small fundamental band gaps for possible applications in infrared photography: The smaller gaps are associated with large average atomic numbers. Hence $(GaSb)_{1-x}Sn_{2x}$, with average atomic numbers ranging from 36.5 to 50, should be an interesting small-band-gap material, provided its electronic transport properties can be made suitable for device applications.

Predicted band gaps of the metastable zinc-blende-diamond Greene alloys fabricated from Al, Ga, In, P, As, Sb, Si, Ge, and Sn are shown in Figs. 2–4. General trends follow those of the prototypical alloy

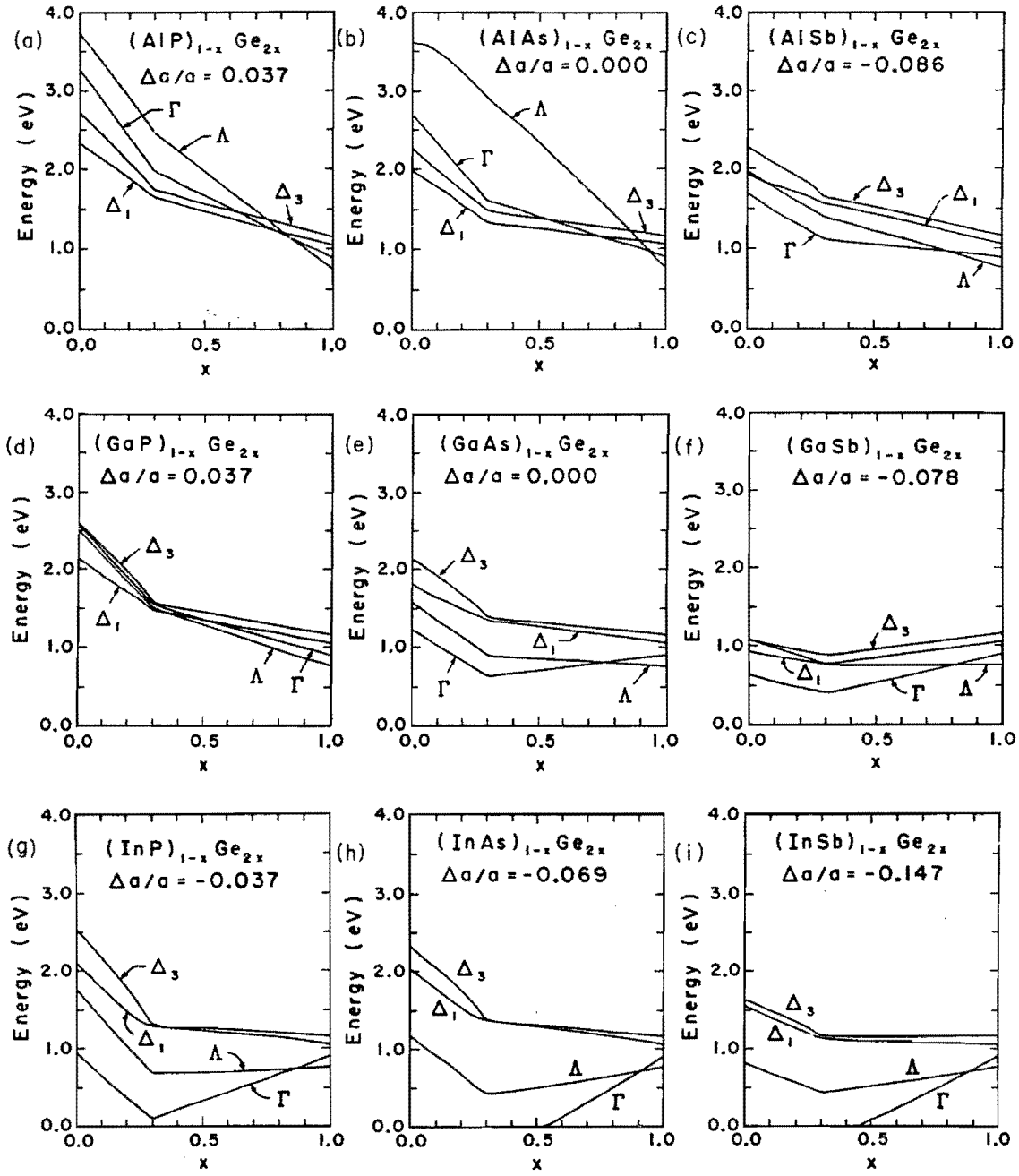


FIG. 3. Predicted band gaps of $(A^{III}B^{V})_{1-x}Ge_{2x}$ alloy versus x, for the following III-V compounds: AlP, AlAs, AlSb, GaP, GaAs, GaSb, InP, InAs, and InSb. Lattice mismatches, defined by Eq. (5), are shown. Kinks can be seen in all levels, at the assumed critical composition $x_c=0.3$. For some alloys, notably $(InP)_{1-x}Ge_{2x}$ for $x<0.4$ and $(InAs)_{1-x}Ge_{2x}$ for $x<0.5$, the Δ minimum occurs at the X point in our tight-binding model and the strict bifurcation at point X is clearly visible.

$(GaSb)_{1-x}Sn_{2x}$, shown in Fig. 1. All alloy band gaps exhibit kinks at x_c as a function of composition. There is always at least one kink in the minimum conduction-band edge at $x=x_c$, due to the phase transition. This kink is not associated with a crossing of the band edges, although these types of effects can also be seen at other compositions. For example, in $(InP)_{1-x}Ge_{2x}$ (Ref. 22) at $x=0.85$, the conduction band at Γ crosses with Λ and the alloy goes from being a direct-gap semiconductor to one with an indirect gap.

The alloys with the smallest lattice mismatches

$$\Delta a/a=(a_{IV}-a_{III\text{-}V})/a_{IV} \tag{5}$$

are especially interesting. We focus primarily on alloys with $\Delta a/a<0.07$. Values of $\Delta a/a$ are given in each figure.

Since the details of the band gaps for these alloys depend on the constituents, we summarize details below fig-

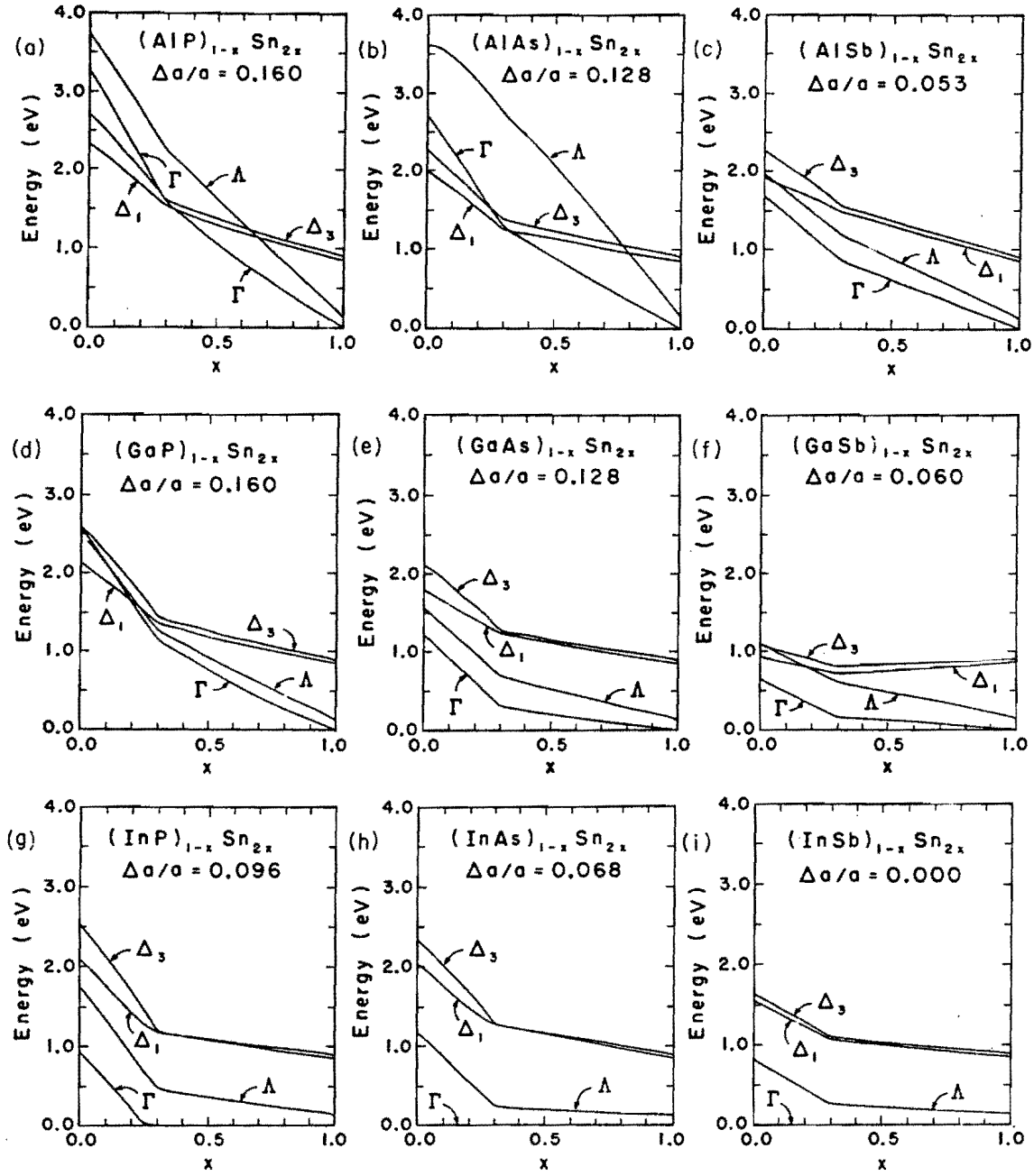


FIG. 4. Predicted band gaps of $(A^{III}B^{V})_{1-x}Sn_{2x}$ alloys versus x, for the following III-V compounds: AlP, AlAs, AlSb, GaP, GaAs, GaSb, InP, InAs, and InSb. Lattice mismatches, defined by Eq. (5), are shown. Kinks can be seen in all levels, at the critical composition $x_c=0.3$. For some alloys, notably $(InP)_{1-x}Sn_{2x}$ for $x<0.6$ and $(InAs)_{1-x}Sn_{2x}$ for $x<0.5$, the Δ minimum occurs at the X point in our tight-binding model and the strict bifurcation at point X is clearly visible.

ure by figure. Figure 2 displays predicted band edges for zinc-blende materials combined in metastable alloys with Si. Those with the smallest lattice mismatches are $(AlP)_{1-x}Si_{2x}$ ($\Delta a/a=-0.004$), $(AlAs)_{1-x}Si_{2x}$ (-0.043), $(GaP)_{1-x}Si_{2x}$ (-0.004), and $(GaAs)_{1-x}Si_{2x}$ (-0.043).[23] Thus, of this class of well-lattice-matched alloys, one is restricted to materials with $\langle Z\rangle \leq 23$. The fundamental band gaps of these alloys vary from 1.17 eV for Si to 2.5 eV for ordinary AlP.[23] These gaps tend to have only one kink, at the critical composition $x=x_c$, because the fundamental gap, like that of Si, is along the Δ_1 line for all x, and does not cross Γ or Λ [the exception being $(GaAs)_{1-x}Si_{2x}$ for which we find crossings from Γ to Λ to Δ as a function of increasing composition]. The kink in Γ for $x\simeq 0.8$ is due to mixing of this level[23] and a Γ_c^{15} level not displayed (Si has $\Gamma_c^{15}<\Gamma_c^1$). In contrast to the small-lattice-mismatched materials, the heavily strained alloys (see the last row of Fig. 2), all show multiple band-

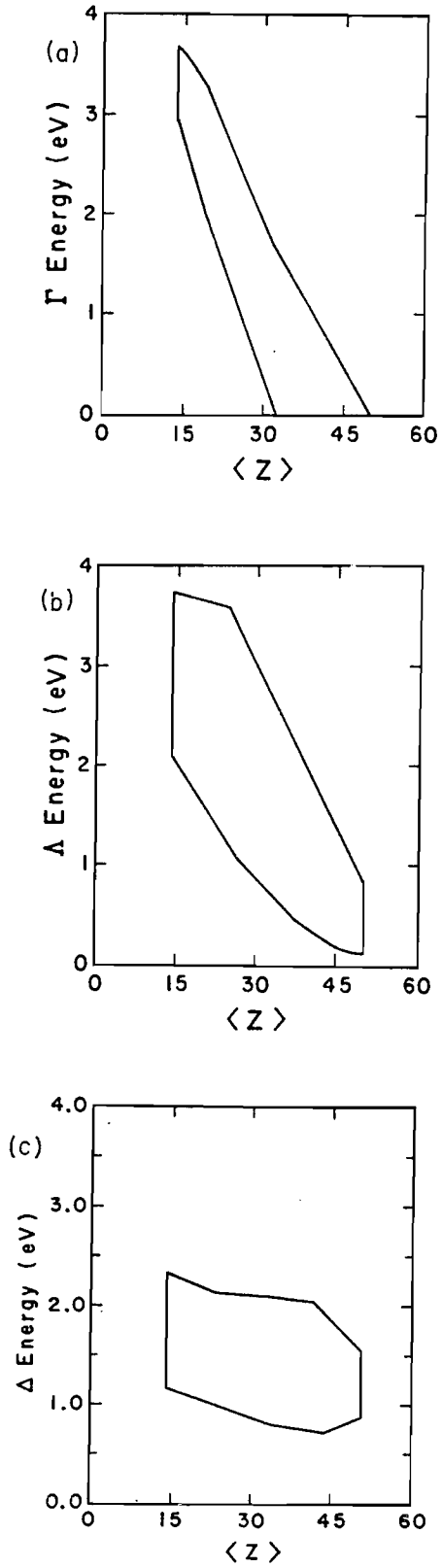


FIG. 5. Trends of the (a) Γ, (b) Λ, and (c) Δ band edges versus average atomic number $\langle Z \rangle$. The relevant energies for the $(A^{III}B^{V})_{1-x}X^{IV}_{2x}$ alloys in question lie within the boxes of the figures. Hence those at Γ and Λ, in particular, tend to decrease with increasing $\langle Z \rangle$.

edge crossings from Γ to Λ to Δ as a function of increasing composition x.

Figure 3 gives band edges for zinc-blende materials in metastable mixtures with Ge. Those with the smallest lattice mismatches are $(AlP)_{1-x}Ge_{2x}$ ($\Delta a/a = 0.037$), $(AlAs)_{1-x}Ge_{2x}$ (0.0), $(GaP)_{1-x}Ge_{2x}$ (0.037), $(GaAs)_{1-x}Ge_{2x}$ (0.0), and $(InP)_{1-x}Ge_{2x}$ (−0.037). In this class of alloys we are restricted to well-lattice-matched materials with $\langle Z \rangle \leq 32$. The band gaps of these alloys vary from 0.1 eV for $(InP)_{1-x}Ge_{2x}$ at $x = 0.3$ to 2.5 for ordinary AlP. The band gaps of these alloys have crossings from Δ to Γ to Λ for $(AlP)_{1-x}Ge_{2x}$ and $(AlAs)_{1-x}Ge_{2x}$ and from Γ to Λ for the others. Of the remaining alloys with larger mismatches, some, such as $(InAs)_{1-x}Ge_{2x}$ and $(InSb)_{1-x}Ge_{2x}$, have zero gap for some compositions x but, because the mismatch is larger, they may be difficult to grow.

Figure 4 presents our predictions for metastable alloys resulting from mixing zinc-blende materials with Sn. Those with the smallest lattice mismatches are $(AlSb)_{1-x}Sn_{2x}$ ($\Delta a/a = 0.053$), $(InAs)_{1-x}Sn_{2x}$ (0.068), $(GaSb)_{1-x}Sn_{2x}$ (0.060), and $(InSb)_{1-x}Sn_{2x}$ (0.0). In this class of alloys, lattice matching restricts us to materials with $32 \leq \langle Z \rangle \leq 50$. These are especially interesting materials because Sn has a zero band gap. The band gaps are predicted to be zero for the metastable alloys $(InAs)_{1-x}Sn_{2x}$ and $(InSb)_{1-x}Sn_{2x}$ for all compositions (despite the fact that the equilibrium compounds InAs and InSb have nonzero gaps[24]). All of the Sn-based metastable alloys (with small lattice mismatches) mentioned above are either direct-gap or zero-gap materials. $(GaSb)_{1-x}Sn_{2x}$ is particularly interesting, because the predicted gap varies from 0.15 eV to zero over a large range in composition, from 0.3 to 1.0. Hence, the gap is small and may not be too sensitive to fluctuations in local environment. This, along with $(InP)_{1-x}Ge_{2x}$, may be an especially good candidate for an infrared detector.[22] The remaining alloys, while covering a large range in gap size, from 2.5 eV for ordinary AlP to zero for Sn, all have large lattice mismatches, $\Delta a/a > 0.096$, and good-quality, long-lived, metastable samples of these materials may be difficult to grow.

IV. CONCLUSIONS

We have presented predictions of the energy band gaps versus alloy composition x for the Greene alloys: metastable, crystalline, substitutional alloys of III-V compounds and group-IV elemental materials. The band gaps at points Γ and L exhibit kinks and the X points bifurcate as functions of composition x, at a critical value x_c corresponding to the order-disorder transition of Newman *et al.* The V-shaped bowing offers the possibility of band gaps significantly smaller than expected on the basis of the conventional virtual-crystal approximation. Alloys with modest lattice mismatches that are predicted to have small band gaps include $(InP)_{1-x}Ge_{2x}$, $(AlSb)_{1-x}Sn_{2x}$, $(GaSb)_{1-x}Sn_{2x}$, and $(InAs)_{1-x}Sn_{2x}$. Larger band-gap alloys with several potentially interesting level crossings in the band gap include $(AlAs)_{1-x}Ge_{2x}$ and $(GaAs)_{1-x}Si_{2x}$.

ACKNOWLEDGMENTS

We gratefully acknowledge the generous support of the U.S. Army Research Office (Contract No. DAAG29-83-K-0122) and the U.S. Office of Naval Research (Contract Nos. N00014-77-C-0537 and N00014-84-K-0352). We thank J. E. Greene and S. Barnett for stimulating conversations about this work and their alloys.

[1]J. E. Greene, J. Vac. Sci. Technol. B 1, 229 (1983).

[2]See also, Z. I. Alferov, R. S. Vantanyan, V. I. Korol'kov, I. I. Mokan, V. P. Ulin, B. S. Yavich, and A. A. Yakovenko, Fiz. Tekh. Poluprovodn. 16, 887 (1982) [Sov. Phys. Semicond. 16, 567 (1982)].

[3]A. J. Noreika and M. H. Francombe, J. Appl. Phys. 45, 3690 (1974).

[4]P. Duwez, R. H. Williams, and W. Klement, Jr., J. Appl. Phys. 31, 1500 (1960) were, to our knowledge, the first to grow some of these alloys for selected compositions.

[5]V. M. Glazov and V. S. Zemskov, *Physiochemical Principles of Semiconductor Doping* (IPST, Jerusalem, Israel, 1978). See also S. I. Shah, K. C. Cadien, and J. E. Greene, J. Electron. Mater. 11, 53 (1980).

[6]K. C. Cadien, A. H. Eltoukhy, and J. E. Greene, Appl. Phys. Lett. 38, 773 (1981); Vacuum 31, 253 (1981).

[7]S. A. Barnett, M. A. Ray, A. Lastras, B. Kramer, J. E. Greene, P. M. Raccah, and L. L. Ables, Electron. Lett. 18, 891 (1982). Typographical errors in the figures have been corrected in Ref. 8.

[8]K. E. Newman, A. Lastras-Martinez, B. Kramer, S. A. Barnett, M. A. Ray, J. D. Dow, J. E. Greene, and P. M. Raccah, Phys. Rev. Lett. 50, 1466 (1983).

[9]K. E. Newman and J. D. Dow, Phys. Rev. B 27, 7495 (1983).

[10]Equation (A17) of Ref. 9 is (obviously) misprinted and should read

$$4KzQ/(k_bT)=\ln[(1-x-Q)^2-M^2]/[(1-x+Q)^2-M^2)]$$
$$+2\ln[(x-Q)/(x+Q)]\ .$$

[11]S. A. Barnett presented preliminary x-ray diffraction data for $(GaSb)_{1-x}Ge_{2x}$ indicating that this does indeed happen: Bull. Am. Phys. Soc. 29, 203 (1984). See also, S. A. Barnett, B. Kramer, L. T. Romano, S. I. Shah, M. A. Ray, S. Fang, and J. E. Greene, *Layered Structure, Epitaxy, and Interfaces*, edited by J. M. Gibson and L. R. Dawson (North-Holland, Amsterdam, 1984).

[12]The critical composition x_c is bounded above by x_p, the percolation composition. An alternative theoretical description of $(GaAs)_{1-x}Ge_{2x}$ in terms of percolation, was given by M. I. D'yakonov and M. E. Raikn, Fiz. Tekh. Poluprovodn. 16, 890 (1982) [Sov. Phys. Semicond. 16, 570 (1982)]. However, percolation would produce a critical composition x_c near 0.6 not 0.3, as observed for $(GaAs)_{1-x}Ge_{2x}$. Preliminary x-ray diffraction studies (Ref. 11) favor the order-disorder model.

[13]M. Blume, V. J. Emery, and R. B. Griffiths, Phys. Rev. A 4, 1071 (1971).

[14]Generally, neither J nor I is known, but both can be simultaneously eliminated from the equation for $M(x;x_c)$ if the phase-transition composition x_c is known. See Ref. 9.

[15]P. Vogl, H. P. Hjalmarson, and J. D. Dow, J. Phys. Chem. Solids 44, 353 (1983). Some parameters have been refit more accuarately to existing data or fit to more recent data in the case of GaSb [T.-C. Chiang and D. E. Eastman, Phys. Rev. B 22, 2940 (1980)]. These are $V(sa,pc)=V(sc,pa)=4.9617$, $V(s^*a,pc)=V(pa,s^*c)=4.5434$ for Ge; $V(sa,pc)=V(sc,pa)=4.2288$, $V(s^*a,pc)=V(pa,s^*c)=3.9665$ for Sn; $V(sa,pc)=4.2485$, $V(sc,pa)=5.2671$, $E(s^*,a)=8.5014$, $V(s^*a,pc)=4.7525$, $V(pa,s^*c)=4.2547$ for GaAs; $E(s,a)=-7.1256$, $E(p,a)=0.6718$, $E(s,c)=-3.7042$, $E(p,c)=2.7312$, $V(s,s)=-5.9854$, $V(x,x)=1.3546$, $V(x,y)=4.4438$, $V(sa,pc)=5.1693$, $V(sc,pa)=4.4708$, $V(s^*a,pc)=5.1609$, and $V(pa,s^*c)=4.1199$ for GaSb.

[16]We have modified the nearest-neighbor model to include second-neighbor interactions, as discussed for Ge and Si by K. E. Newman and J. D. Dow, Phys. Rev. B 30, 1929 (1984).

[17]W. A. Harrison and S. Ciraci, Phys. Rev. B 10, 1516 (1974); W. A. Harrison, *Electronic Structure and the Properties of Solids* (Freeman, San Francisco, 1980).

[18]L. Vegard, Z. Phys. 5, 17 (1921).

[19]Y. Onodera and Y. Toyozawa, J. Phys. Soc. Jpn. 24, 341 (1968).

[20]H. Holloway and L. C. Davis, Phys. Rev. Lett. 53, 83 (1984), have recently suggested that the satisfaction of the Onodera-Toyozawa criterion may not be sufficient to guarantee the validity of a virtual-crystal approximation. We disagree with many of their statements about Ref. 9. The central theoretical arguments of their paper are contradicted by recent experimental results (Ref. 11).

[21]The relative minimum of the conduction band is either at point Γ, or near (or at) points X or L, in the Δ or Λ direction.

[22]D. W. Jenkins, K. E. Newman, and J. D. Dow, J. Appl. Phys. 55, 3871 (1984).

[23]We are using a notation Γ for s-like states Γ^1, as opposed to p-like states Γ^{15}_c (c denoting the conduction band and v denoting valence band).

[24]Recall that because of ion bombardment during growth, metastable AlP, for example, has more antisite defects and thus should be different from stable equilibrium AlP. This model gives a smaller band gap for metastable III-V compound semiconductors than for the stable III-V compounds. The "magnetization" $M(x;x_c)$ is not unity for $x=0$ in this theory, i.e., the theory predicts a significant concentration of antisite defects.

Applications of Surface Science 22/23 (1985) 937–947
North-Holland, Amsterdam

INTERFACIAL DEEP LEVELS RESPONSIBLE FOR SCHOTTKY BARRIER FORMATION AT SEMICONDUCTOR/METAL CONTACTS

John D. DOW

Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556, USA

Otto F. SANKEY

Department of Physics, Arizona State University, Tempe, Arizona 85287, USA

and

Roland E. ALLEN

Department of Physics, Texas A&M University, College Station, Texas 77843, USA

Received 27 August 1984; accepted for publication 31 October 1984

The following facts, and many others, concerning III–V (e.g., GaAs, InP) Schottky barriers can be understood in terms of Fermi-level pinning by interfacial antisite defects (sheltered by vacancies) at semiconductor/metal contacts: (i) the barrier heights are almost independent of the metal in the contact; (ii) the surface Fermi levels can be pinned at sub-monolayer coverages and the pinning energies are almost unaffected by changes of stoichiometry or crystal structure; (iii) the Schottky barrier height for n-InP with Cu, Ag, or Au is $\simeq 0.5$ eV, but changes to $\simeq 0.1$ eV when reactive metal contacts (Fe, Ni, or Al) are employed because the antisite defects are dominated by P vacancies; and (iv) the dependence on alloy composition for alloys of AlAs, GaAs, GaP, InAs, and GaAs is extremely complex – owing to the dependence of the binding energy for the cation-on-anion-site deep level on alloy composition. Fermi-level pinning by Si dangling bonds at Si/transition-metal silicide interfaces accounts for the following facts: (i) the barrier heights are independent of the transition-metal, to within $\simeq 0.3$ eV; (ii) on the 0.1 eV scale there are chemical trends in barrier heights for n-Si, with the heights decreasing in the order Pt, Pd, and Ni; (iii) barriers form at low metallic coverage, (iv) barrier heights are independent of silicide crystal structure or stoichiometry to ± 0.1 eV; and (v) the barrier heights for n-Si and p-Si add up to approximately the energy of the band gap.

1. Introduction

When a metal is deposited on a semiconductor surface, a potential barrier to electron motion is formed, which prevents the flow of electrons between the metal and the semiconductor. The physics governing the formation of this Schottky barrier is controversial even today. Here we present theoreti-

cal calculations which support the notion that Schottky barriers are commonly (but not exclusively [1]) formed as a result of "Fermi-level pinning" by deep trap states associated with defects at the semiconductor surface.

The basic idea of Fermi-level pinning was enunciated by Bardeen [2] in 1947, and is most easily described for the limit of a degenerate n-type semiconductor in contact with a metal. The Fermi energies of the semiconductor, the semiconductor surface, and the metal must align in electronic equilibrium. At zero temperature, the Fermi level of the semiconductor lies almost at the conduction band edge (more precisely, at the donor level), and lines up with the Fermi level of the metal. The Fermi energy of the semiconductor's surface, however, can lie deep in the fundamental band gap if there are deep impurity levels in the gap. In this case of sufficient concentration of deep levels in the gap, the deep levels determine and "pin" the Fermi energy of the surface, which does not align with the bulk semiconductor's Fermi energy if the valence band maxima of the bulk and the surface are assumed to be at the same energy. Hence, the semiconductor and its surface are not in electronic equilibrium when the valence band maxima align. As a result, carriers must diffuse in order to bring the surface into electronic equilibrium with the bulk semiconductor and the metal: a surface dipole must build up, and the bands must bend near the surface to align the Fermi energies of the bulk and the surface. This results in a Schottky barrier (see fig. 1). Bardeen, in his Fermi-level pinning paper, left open the possibility that the deep levels responsible for the pinning might be either intrinsic (e.g., surface states) or extrinsic. Spicer and co-workers [3] have championed the idea that *native defects* produced during the formation of the semiconductor/metal contact pin the Fermi energy.

In this Fermi-level pinning model, one can estimate the Schottky barrier height for an n-type semiconductor by first determining the defect responsible for the pinning and then calculating the difference in energy between

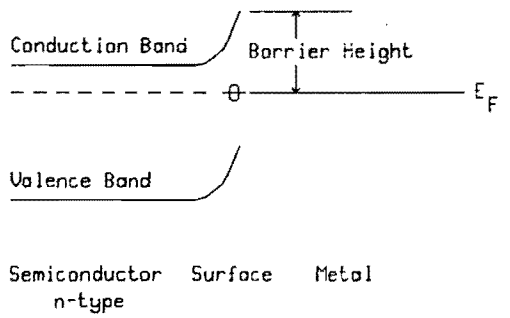


Fig. 1. Schematic illustration of the Bardeen model of Fermi-level pinning by a semiconductor-surface defect deep level (denoted by an open circle with a bar through it).

the neutral defect's lowest unfilled deep level in the gap and the conduction band edge. (For p-type material, the highest filled level in the gap pins the Fermi level.) Hence the problem of calculating the Schottky barrier height is reduced to the equivalent problem of computing the binding energy of a deep level. (See fig. 1.)

2. Deep levels in the bulk

To understand the physics of deep levels at metal/semiconductor contracts, one must first comprehend the basic physics of deep impurities in the bulk of a semiconductor. An impurity level, by current definition [4], is "deep" if that level originates from the central-cell defect potential of the impurity (as opposed to originating from the long-ranged Coulombic tail of the defect potential, as for "shallow" levels). In covalently bonded semiconductors, sp^3-bonded substitutional defects have typically *four* deep levels near the fundamental band gap and an infinite number of shallow levels. The infinite number of shallow levels is associated with the fact that the Coulomb potential has an infinite number of bound states, and the four deep levels are due to there being one s-like and three p-like orbitals for an sp^3-bonded defect. In the bulk of a tetrahedral semiconductor, the three p-states are degenerate, forming a T_2-symmetric deep level, and the s-state gives rise to an A_1 level.

The four deep levels need not all lie within the fundamental band gap, however. In fact, it is rare that all four do. Indeed, a "shallow impurity" is one for which all of its deep levels lie *outside* the fundamental band gap (fig. 2). A "deep impurity" is an impurity that produces at least one deep level in the band gap. The issue of whether a deep level lies within the gap or not is a quantitative one: if the host bands are broad enough and the fundamental band gap is narrow enough, then the bands are likely to cover up all of the deep levels, making them resonant with the host bands. Hence, narrow-gap semiconductors tend to have relatively fewer "deep impurity" centers (with levels in the gap) than large band-gap materials.

The basic physics of deep levels is illustrated schematically in fig. 3 for the case of an N impurity replacing P in the bulk of GaP. For simplicity we consider only the A_1 or s-like deep state of the defect. First consider atomic Ga and P, which, when combined into a molecule, form bonding and antibonding levels. The bonding–antibonding splitting is of order $v^2/(\epsilon_{Ga} - \epsilon_P)$, where v is the nearest-neighbor transfer matrix element and $\epsilon_{Ga} - \epsilon_P$ is the energy denominator resulting from perturbation about the extreme tight-binding limit [5,6]. The bonding and antibonding states of the molecule are the parents of the conduction and the valence bands of the solid, respectively. If now one P atom is replaced by an N impurity atom, the N

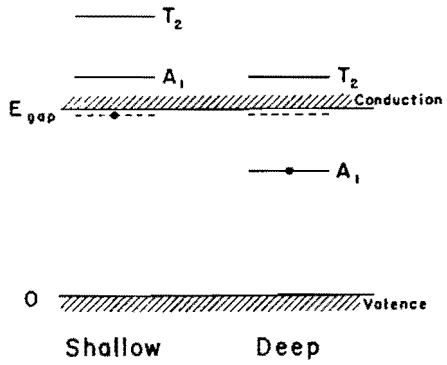


Fig. 2. Schematic illustration of the difference between "shallow" and "deep" sp^3-bonded substitutional (donor) impurities, after ref. [21]. The shallow energy levels in the band gap are dashed. The deep levels of A_1 (s-like) and T_2 (p-like) symmetry are denoted by heavy lines. In the case of a "shallow impurity" the deep levels are resonances and lie outside the fundamental band gap; for a "deep impurity" at least one deep level lies within the gap. The lowest level is occupied by an extra electron (dark circles) if the impurity has a valence one greater than the host atom it replaces (e.g., S or O on a P site in GaP).

will try to hybridize with its neighbors. However, the atomic energy of the N is $\simeq 7$ eV lower than the corresponding energy of the P atom it replaces (i.e., the defect potential in the central-cell is $V \simeq -7$ eV). As a result, the energy denominator is $\simeq 7$ eV larger for N than for P, and (since v is almost the same for P and N [7]), the bonding–antibonding splitting is smaller – and the deep level lies within the band gap. For a slightly less negative value of V (i.e., a slightly more electropositive defect than N, such as S), however, the deep level is resonant with the conduction band – so that at most "shallow" states bound by the long-ranged Coulombic, $-Ze^2/\epsilon r$, part of the defect potential (neglected here) would lie in the gap. (Here Z is the impurity-host valence difference and is zero and unity for N and S, respectively, replacing a P atom in GaP.)

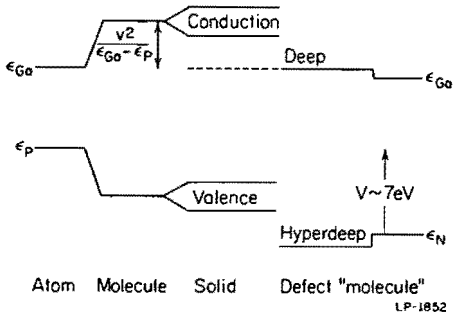


Fig. 3. Schematic illustration of the qualitative physics of deep levels, as discussed in ref. [4].

3. Deep levels of surface defects

The same basic physics holds for a defect at a surface. The reduction of the tetrahedral symmetry by the surface causes the A_1 bulk levels to shift and the T_2 deep levels to split into three orbitally non-degenerate levels. For a defect at a free surface the splittings of the T_2 level are of order 1 eV. Therefore, the free-*surface defect levels lie at substantially different energies from the bulk defect levels.*

A central question is whether the pinning defects for Schottky barrier formation are at the semiconductor's surface or in the semiconductor's bulk. There is no definitive experimental answer to this question at the present, but we believe that the defects are *at or very near* the semiconductor/metal interface *in environments that are quite similar to the environment of a defect at a free surface.* By this we mean the pinning defects are each adjacent to a vacancy or a void (or a highly electropositive atom) that "shelters" it electronically from its more distant neighbors: a defect–vacancy pair in the semiconductor's bulk has essentially the same energy levels as a defect at a surface [8] – because deep-level wavefunctions are rather localized to the shell of first-neighbors of the defect, and the main difference between a defect–vacancy pair and the same defect at a surface is that, at the surface (which can be thought of as a sheet of vacancies), some second- and more-distant neighbors are vacancies rather than atoms. (Second-neighbor effects on a deep level are rarely major.)

Our reasons for adopting this viewpoint that the pinning defects are near the semiconductor/metal interface and "sheltered" in free-surface-like environments are: (i) Fermi-level pinning can occur at sub-monolayer metallic coverages, a fact that is difficult to explain unless the relevant defects are at or near the interface; (ii) the simple bulk point defects, such as vacancies and antisites, unquestionably give qualitatively as well as quantitatively incorrect predictions for the observed behavior of Fermi-level pinning and Schottky barrier heights (e.g., the bulk antisite As_{Ga} – As on a Ga site – in GaAs cannot explain the Fermi-level pinning for n-GaAs because it produces only an occupied deep donor level in the gap, whereas an unoccupied acceptor is required to achieve Fermi-level pinning in the gap for n-type material – i.e., the next available level for an electron is the deep level, rather than the conduction band edge). In contrast, As_{Ga} at the surface produces *two* deep levels in the gap: a deep donor and a deep acceptor; (iii) without the concept of sheltering, the defect theory would be in conflict with the experimental fact that, for GaAs and some other semiconductors, the deposition of different (non-reactive) metals in a semiconductor/metal contact most often leads to the same Schottky barrier height (if the defect were in direct contact with the metal, its energy levels would be significantly altered by changing the metal). Therefore, the pinning defect must be

adjacent either to a vacancy or to a very electropositive atom (recall that electronically a vacancy is an infinitely electropositive "atom" [9]).

Spicer and co-workers [3], Wieder et al. [10], Mönch et al. [11], and Williams et al. [12] have presented numerous data which indicate that the surfaces of III–V semiconductors have Fermi levels determined or "pinned" by the deep impurity states of *native defects*. The exact mechanisms by which these defects are created are not presently understood, but it is believed that they are normally generated during the formation of the surface (e.g., by cleavage) or during the deposition of a metal contact. Indeed, the precise nature of the native defects is not presently known, and one purpose of this work is to provide a theoretical framework for identifying the "pinning defects". We shall enumerate the possible native defects, argue that the pinning levels of many complex defects are virtually identical to the pinning levels of a few simple ones, show that some simple defects can explain the observed chemical trends in Schottky barrier data for III–V semiconductors while others cannot, and propose a relatively simple and specific picture of the pinning defects.

The possible native defects are anion and cation vacancies, both types of antisite defects, anion and cation interstitials, and combinations of these. It can be shown, however, that the combination defects normally have spectra similar to the sum of their constituents' spectra [8] – and so we consider only the isolated defects. We also eliminate interstitials from consideration, because (i) interstitials are known to be very sensitive to the local environment [13] (whereas Fermi-level pinning defects are not), and (ii) in the bulk, the Group-III and Group-V atoms have been observed either on their own sites or on the antisite, but (to our knowledge) not at interstitial positions. Defects associated with the metal atoms originating from the metal of the contact are not considered because (i) for some semiconductors at least, the Schottky barrier heights are relatively independent of the metal, and (ii) for most of the semiconductors of interest, the metal atoms themselves do not produce the required deep levels in the fundamental band gap.

Thus we are left with an apparently simple problem: compute the deep levels of the vacancies and the antisite defects, and determine if these levels explain the observations. In making these calculations, however, we must recognize that this or any theory has uncertainties of order ≈ 0.5 eV (part of which is due to the neglect of lattice relaxation around the defect). Therefore, we do not simply compare the theory with data, but instead we (i) eliminate as many as possible of the Fermi-level pinning assignments because the theory and the data disagree by > 0.5 eV, and (ii) make our final assignments on the basis of the observed chemical trends in the Fermi-level pinning positions from one semiconductor to another.

The calculations employ an empirical tight-binding Hamiltonian [6] for the host semiconductor. Since the parameters of this Hamiltonian exhibit

chemical trends, the defect potential's matrix elements can be estimated from the trends [4]. In the localized-orbital sp^3s^* tight-binding basis, the defect potential V is diagonal (provided we assume that the lattice does not relax around the defect) with elements proportional to the differences between the atomic energies of the defect and the host atom it replaces. The deep levels of the defect are obtained by solving the secular equation

$$\det[1 - (E - H_0)^{-1}V] = 0 \ ,$$

where H_0 is the host Hamiltonian and E is the deep level. For the calculations reported here, H_0 describes a relaxed (110) III–V surface with the 27° rigid rotation of the anions out of the surface plane, and V is a matrix simulating the central-cell potential of a defect at the surface. However, a far simpler model involving defect–vacancy pairs in the bulk or at an interface would give similar results [14]: in the simpler case H_0 would represent the sheltering vacancy and the cluster of atoms at the defect site (before the defect is introduced) and at surrounding first- and possibly more-distant-neighbor sites. The details of solving the secular equation either for a defect at the free surface [15] or for a simplified cluster model [14] have been described elsewhere.

4. Results for III–V semiconductors

The results of our calculations of the Schottky barrier heights (i.e., the binding energies of the lowest incompletely occupied one-electron level of the neutral impurity with respect to the conduction band edge) are given in fig. 4, where we have assumed that the defect responsible for Fermi-level pinning is the cation-on-anion-site antisite defect at the surface. The agreement between theory and data is strikingly excellent, and strongly supports the hypothesis that this antisite defect is responsible for the observed Schottky barrier formation. (The two vacancies and the other antisite defect fail to reproduce all the observed trends.)

This success does not mean that all Schottky barrier formation in III–V semiconductors is attributable to Fermi-level pinning by cation-on-anion-site defects. Although an antisite defect can be formed with less free energy that a vacancy [16], we believe Fermi-level pinning by vacancies has been observed for InP contacts with reactive metals [17]. Indeed, the apparent dependence of Schottky barrier height on chemical reactivity [12,17,18] can be explained in terms of chemical reactions changing the dominant defect from an antisite to a vacancy. The reactive metals combine with P making stable compounds, leave P vacancies (V_P). In InP these vacancies are predicted to yield shallow donor levels in the fundamental gap near the conduction band edge; these levels pin the Fermi energy and yield a small

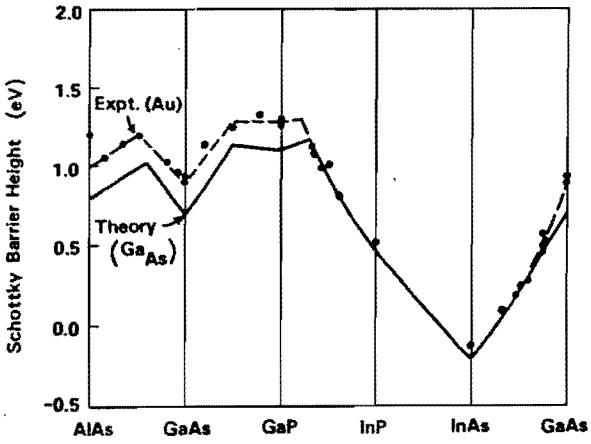


Fig. 4. Experimental (dashed) and theoretical (solid) dependence of the Schottky barrier heights of III–V semiconductor alloys with Au contacts versus alloy composition, after ref. [22]. The theory assumes Fermi-level pinning by a surface cation-on-anion-site defect.

(≃ 0.1 eV) Schottky barrier height. Thus, when reactive metals (viz., Fe, Ni, or Al) are deposited on InP, the dominant Fermi-level pinning defects appear to be P vacancies; but when non-reactive metals are deposited (viz., Cu, Ag, or Au), the antisite defect levels appear to dominate, and the barrier height is approximately 0.5 eV. Thus the theory, supplemented by the hypothesis that the reactive metals produce P vacancies, can account for the InP data.

The ability of the theory to provide a natural explanation of the dependence of Schottky barrier height on chemical reactivity is especially important, because it offers a resolution of a major controversy between the viewpoints (i) that Schottky barrier formation is due to Fermi-level pinning by defects (championed by Spicer and co-workers [3]), and (ii) that Schottky barrier formation depends critically on chemical reactivity (advocated by Brillson and associates [18]). Our own viewpoint is that both sides of the controversy are essentially correct, and that different chemical reactions produce different dominant defects and Fermi-level pinning positions.

Presently, it is not known if reactive metals do indeed produce a sufficient number of interfacial P vacancies in InP. Indeed, studies of P diffusion indicate that the diffusion rate is greater for non-reactive metals and that P concentrates at the reactive-metal/InP interface [19]!

There are many other experimental facts concerning III–V semiconductors, most of them of a detailed nature, that the Fermi-level pinning theory can explain. But rather than focus on those details in our limited space, we instead turn our attention to Si (which, being homopolar, has no antisite defects) and the question of whether Si's Schottky barriers are similar to those of the III–V's.

5. Si/transition-metal silicide Schottky barriers

The best-studied Si Schottky barriers are those with transition-metal silicides. The silicides themselves are metals created by the reaction of the transition-metals with the Si. For these systems we believe the pinning defects are Si dangling bonds.

The local defect we think is responsible for the Fermi-level pinning is similar to a bulk-Si vacancy whose four nearest-neighbor atoms (instead of all being Si) are three transition-metal atoms and one Si atoms (the one whose bond dangles into the sheltering vacancy due to a missing Si bridge atom at the Si/transition-metal silicide interface) [20]. If the transition-metal atoms were instead Si atoms, locally this defect would be a bulk-Si vacancy – with an A_1 symmetric deep level resonant with the valence band and a T_2 level in the Si band gap. Hence, to determine the physics of the Fermi-level pinning at the Si/transition-metal silicide interface, we need only understand how the bulk-Si vacancy's deep levels change as three of the vacancy's neighbors change from Si into transition-metal atoms. The change of the three neighbors from Si into transition-metal atoms can be simulated by increasing the sp^3 hybrid energies of the atoms on the transition-metal sites (for the hybrids oriented toward the vacancy) from ϵ_h (for Si) to $\epsilon_h + V$ (for transition-metal atoms), with V of order 5 eV. That is, relative to Si, the transition-metal atoms are very electropositive (electronically like vacancies). The large positive repulsive potential V on the transition-metal sites merely pushes the Si dangling bond away from the silicide and into the Si. In the process, it drives the energy of the T_2 deep level for the bulk-Si vacancy out of the fundamental band gap and into the conduction band, and brings the A_1 level up into the gap. (For $V \to +\infty$, the A_1-derived level approaches the hybrid energy ϵ_h asymptotically from below.) This level, for the neutral defect, is singly occupied by one electron, and therefore can pin the Fermi energy of either n-Si or p-Si. Hence the barrier heights for n- and p-type material add up to the band gap. Since changes of V of order 1 eV have little effect on a pinning level that asymptotically approaches ϵ_h, the theory explains why different transition-metals have the same barrier heights to within $\simeq 0.3$ eV, while the differences in barrier heights on the 0.1 eV scale reflect the chemical trends in V (which is proportional to the difference between the atomic energy of a transition-metal and that of Si) giving decreasing barrier heights for Pt, Pd, and Ni silicides. Moreover, since the pinning defect is localized and has properties that depend primarily on the electropositivity of the transition-metal atoms, one can understand why barriers form at low metallic coverages and have heights that are insensitive to stoichiometry or the silicide crystal structure. Thus Fermi-level pinning by dangling bonds can account for the main experimental facts concerning Schottky barrier heights at Si/transition-metal silicide interfaces.

6. Discussion

We have touched on a small subset of the many and varied experimental facts that can be explained by the Fermi-level pinning model of Schottky barrier formation. This model has been successfully applied to Schottky barriers involving Si, $Si_{1-x}Ge_x$, and diamond (with Fermi-level pinning by dangling bonds) as well as to III–V semiconductors (in which antisite defects and dangling bonds pin the Fermi level). It appears to be applicable to any covalent semiconductor which responds to contact formation by spontaneously producing a sufficient number of native defects.

Acknowledgements

We are grateful to the US Office of Naval Research and the Army Research Office for supporting this research (Contract Nos. N00014-84-K-0352 and DAAG29-83-K-0122). We also thank D. Jenkins for his assistance with the figures.

References

[1] Materials like GaSe that are resistant to the formation of surface defects do not exhibit Fermi-level pinning, but obey the Schottky model.
[2] J. Bardeen, Phys. Rev. 71 (1947) 717.
[3] W.E. Spicer, P.W. Chye, P.R. Skeath, C.Y. Su and I. Lindau, J. Vacuum Sci. Technol. 16 (1979) 1422;
W.E. Spicer, I. Lindau, P.R. Skeath and C.Y. Su, J. Vacuum Sci. Technol. 17 (1980) 1019.
[4] H.P. Hjalmarson, P. Vogl, D.J. Wolford and J.D. Dow, Phys. Rev. Letters 44 (1980) 810;
H.P. Hjalmarson, PhD Thesis, University of Illinois (1979) (unpublished);
H.P. Hjalmarson, P. Vogl, D.J. Wolford and J.D. Dow (unpublished).
Many of the ideas of this work are based on earlier work by W.Y. Hsu, J.D. Dow, D.J. Wolford and B.G. Streetman, Phys. Rev. B16 (1977) 1597.
[5] ϵ_{Ga} and ϵ_P here are s atomic orbital energies, as given in ref. [6].
[6] P. Vogl, H.P. Hjalmarson and J.D. Dow, J. Phys. Chem. Solids 44 (1983) 365.
[7] W.A. Harrison, Electronic Structure and the Properties of Solids (Freeman, San Francisco, 1980).
[8] O.F. Sankey and J.D. Dow, Phys. Rev. B26 (1982) 3243, and references therein.
[9] M. Lannoo and P. Lenglart, J. Phys. Chem. Solids 30 (1969) 2409.
[10] H.H. Wieder, Surface Sci. 132 (1983) 390, and references therein.
[11] W. Mönch, Surface Sci. 132 (1983) 92, and references therein.
[12] R.H. Williams, Surface Sci. 132 (1983) 122, and references therein.
[13] O.F. Sankey and J.D. Dow, Phys. Rev. B27 (1983) 7641.
[14] O.F. Sankey, R.E. Allen and J.D. Dow, to be published.
[15] J.D. Dow, R.E. Allen and O.F. Sankey, in: Chemistry and Physics of Solid Surfaces, Vol. 5, Eds. R. Vanselow and R. Howe (Springer, Berlin, 1984).

[16] J.A. Van Vechten, J. Electrochem. Soc. 122 (1975) 419, 423. The defect formation process is almost certainly an event that occurs far from equilibrium.
[17] J.D. Dow and R.E. Allen, J. Vacuum Sci. Technol. 20 (1982) 659.
[18] L.J. Brillson, Surface Sci. Rept. 2 (1982) 132, and references therein.
[19] Y. Shapira and L.J. Brillson, J. Vacuum Sci. Technol. B1 (1983) 618.
[20] O.F. Sankey, R.E. Allen and J.D. Dow, Solid State Commun. 49 (1984) 1.
[21] J.D. Dow, Lectures (Varenna), to be published.
[22] J.D. Dow and R.E. Allen, to be published.

Applications of Surface Science 22/23 (1985) 948–968
North-Holland, Amsterdam

ADVANCES IN CHARACTERIZING AND CONTROLLING METAL–SEMICONDUCTOR INTERFACES *

L.J. BRILLSON

Xerox Webster Research Center, 800 Phillips Road, W-114, Webster, New York 14580, USA

Received 27 August 1984; accepted for publication 7 November 1984

We have used a variety of novel approaches in characterizing metal–semiconductor interfaces – soft X-ray photoemission spectroscopy with interlayers or markers, surface photovoltage spectroscopy, and cathodoluminescence spectroscopy, coupled with pulsed laser annealing – to reveal systematics between interface chemical and electronic structure. The chemical basis for these interfacial properties suggests new avenues for controlling electronic structure on a microscopic scale.

1. Introduction

With the application of surface science techniques to the study of metal–semiconductor interfaces, considerable progress has been achieved in understanding the interactions which take place at the microscopic junction and their influence on macroscopic electronic properties [1–6]. In particular, it is now generally accepted that the extrinsic electronic states of a metal–semiconductor interface – e.g., those due to some interaction between metal and semiconductor – rather than any intrinsic states present at the semiconductor surface – dominate the Schottky barrier formation. Considerable evidence for these conclusions has been derived from contact potential [7,8], surface photovoltage, low energy electron loss [9,10], UV [12,13], and soft X-ray photoemission spectroscopies [14–21]. With these techniques, research groups around the world have found strong charge transfer and atomic redistribution occurring with the deposition of only a few monolayers or less of deposited metal on clean, ordered semiconductor surfaces. Thus related phenomena such as chemical reactions, diffusion, formation of defects, dipoles, and alloy layers at the metal–semiconductor interface are observed which can account for Schottky barrier formation on an atomic scale. Within the last few years, this body of work has been extended to reveal further

* This work reported here was carried out in collaboration with C.F. Brucker, A. Katnani, M. Kelly, G. Margaritondo, H. Richter, Y. Shapira, M. Slade, and N.G. Stoffel.

0378-5963/85/$03.30

Scanning tunneling microscope tip structures

Ruth Nicolaides
U.S. Army ARDEC, Dover, New Jersey 07801

Yong Liang,[a] William E. Packard,[a] Zhou-Wu Fu,[a] Howard A. Blackstead,[a] K. K. Chin,[a] John D. Dow,[a] Jacek K. Furdyna,[a] Wei Min Hu,[a] Robert C. Jaklevic,[a),b)] William J. Kaiser,[a),c)] Alan R. Pelton,[d)] and Mary V. Zeller[e)]
University of Notre Dame, Notre Dame, Indiana 46556

Joseph Bellina, Jr.
Department of Physics and Chemistry, St. Mary's College, Notre Dame, Indiana 46556

(Received 24 September 1987; accepted 29 October 1987)

Studies of electrochemically etched tungsten scanning tunneling microscope tips, using scanning electron microscopy, show that (i) the tips are often convolved or bent if the mass of the tungsten wire submerged in the etchant is large (an effect ascribed to surface plastic flow), (ii) bent tips nevertheless often produce good quality scanning tunneling microscopy images of Au films in air, but (iii) tips, once crashed clumsily into the Au films, no longer produce images.

I. INTRODUCTION

In an ideal scanning tunneling microscope, electron tunneling occurs between the surface being studied and a single atom at the end of a sharp tunneling tip. In practice, it is rare to prepare tips even resembling the sharp, single-atom ideal. Often, in order to obtain scanning tunneling microscope images from layered compounds, one must first crash the tungsten tip into the surface. This initial crash very likely "spears" a layer of the material being studied, which then can act as a tunneling tip. For example, scanning tunneling microscope studies of layered compounds, such as graphite, using tungsten tips suggest that the "tip" may in reality be a layer of graphite stuck on the tungsten.[1,2] Colton *et al.*[1] and Mizes and Harrison[3] have shown rather dramatically that many of the different images reported for graphite surfaces can be obtained by having more than one atom acting as a tunneling site. In III–V semiconductors, Feenstra and Fein[4] have shown that images of defects on the GaAs (110) surface depend on the character of the tip as much as on the defect. Biegelsen *et al.*[5] have published studies of tip structures and have found that ion milling improves the sharpness of a tip, removes oxide, and enhances the tip's reliability. Clearly the role of the tip and its geometry in forming scanning tunneling microscope images is incompletely understood.

In this paper, we report some elementary studies of tungsten scanning tunneling microscope tips. These include studies of scanning electron microscope images of tips, the dependence of tip geometry on tip etching and growth conditions, and the quality of scanning tunneling microscope images obtained from each tip. As our touchstone of comparison, we use images of Au films in air. Surface Au atoms have a high mobility, forming nearly planar surfaces, and the steps on these surfaces are easily visible with our microscope. We use Au rather than graphite as our standard because graphite layers are too easily peeled from the surface. We find, not surprisingly, that once our tips crash into the surface of Au, unless the crash is rather gentle,[6] the tips no longer produce good images; however, we also find that tip geometry, as observed with a scanning electron microscope, can be a deceptive predictor of scanning tunneling microscope image quality. In particular, some tips can be terribly "bent" or convolved geometrically and yet produce rather good images.

II. TIP PREPARATION

Each tip was prepared by placing several millimeters of the lower end of a tungsten wire (0.025 in. diameter) into an aqueous 1M NaOH etching solution and applying a 12-V potential to the tungsten wire (with respect to a stainless-steel electrode inserted into the solution). The etch was continued until the submerged portion of the wire dropped off into the bath, leaving the usable tip suspended near the liquid/air interface. By electronically monitoring the etching current (typically 10 mA) with a comparator circuit, the 12-V potential was shut off when the wire separated. This prevented further etching of the tip. After the separation, the etch voltage was pulsed "on" for 1 s, to remove any irregularities at the end of the tip.

In order to prevent unnecessary etching, we covered a large portion of the wire submerged in the solution with Teflon insulation. This kept the current density in the etching region approximately constant and permitted better determination of the mass of the submerged portion of the wire (for correlation of tip shape with the mass of the submerged portion, see below).

III. SCANNING ELECTRON MICROSCOPE IMAGES

The tips that we etched generally exhibited nearly exponential shapes (Fig. 1), rather than the nearly paraboloic shapes reported by some authors.[7] We find that this exponential shape results when the current density (and hence the reaction rate) is high. We have observed that, with a longer length of wire (~ 1 cm or more) exposed to the etchant, the profile of the tip tended to become more parabolic. We have also found that more paraboiic shapes result from electrochemical etching with alternating rather than direct current.

A number of our tips had, in addition to the nearly expo-

 0734-2101/88/020445-03$01.00

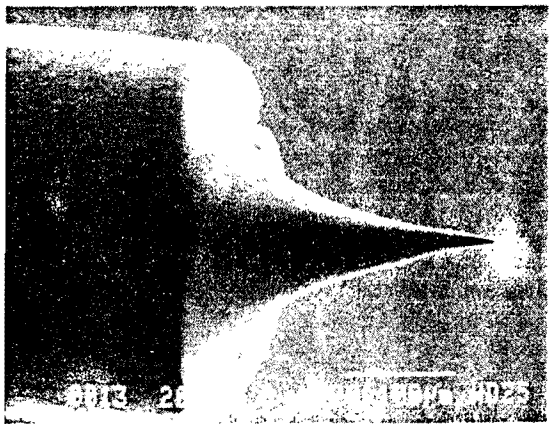


FIG. 1. Scanning electron microscope image of a pointed tungsten tip. The tip was etched in an aqueous NaOH solution. Note the 100-μm scale.

nential overall shape, highly convolved or bent points (Fig. 2), although such tips had never been allowed contact with any surface. By carefully controlling the etching conditions, we learned that such bent points tend to occur when the mass (length) of the tungsten wire in the etchant is large, a condition indicative of (i) plastic flow of the tungsten wire as the tip is formed and (ii) some recoil of the tip at the instant of tip formation, when the wire in the etchant drops off.

Just before the bottom portion of the wire separates, plastic flow occurs at the narrowest region of the wire when the stress induced by the wire's weight is greater than the yield stress. Rough estimates indicate that the weight of several millimeters of tungsten wire in the etchant bath is sufficient to allow plastic deformation at a necking diameter of about 1 μm. Furthermore, the mechanical energy stored in the neck region of the stretched tungsten wire is released when the wire separates. This energy, although perhaps an order of magnitude too small to plastically deform the entire volume of the thicker portion of the wire, is nevertheless sufficient to deform small surface regions, leading to tip recoil and bending, with the yield stress apparently being exceeded locally at certain surface regions.

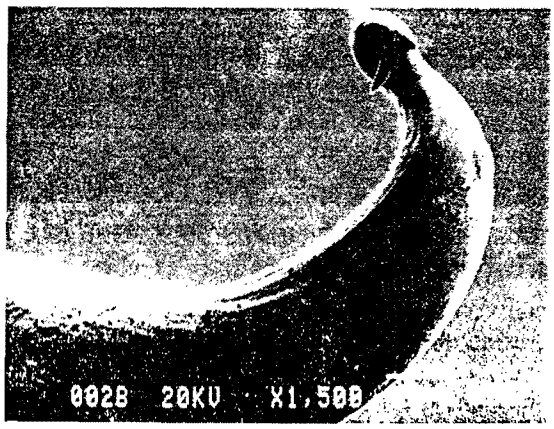


FIG. 2. Scanning electron microscope image of a contorted or bent tungsten tip. The contortion is not due to the tip's having been crashed, but rather is due to tip deformation during etching. Such bent tips occur when a large mass of wire is submerged in the etchant, and are ascribed to recoil after fracture resulting from plastic deformation. Note the 10-μm scale. Note also the "dirt" on the tip, residual NaOH.

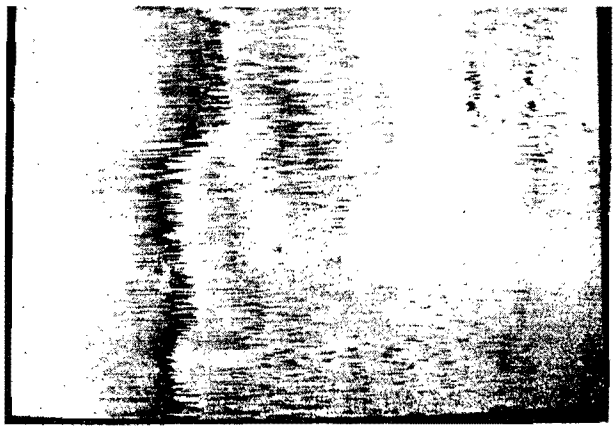


FIG. 3. Scanning tunneling microscope image of a Au film, taken using the tip of Fig. 1. This is a tunneling current image over a 150 × 150 Å area of the film. Comparable quality images are obtained with both pointed and bent tips, provided the tips have not been crashed.

The possibility of plastic flow playing a role in the formation of tips has been raised previously by Müller and Tsong,[8] but those authors ascribed the tip bending to the action of gas bubbles. We virtually eliminate such bubbles by using a direct-current etch (alternating current produces many bubbles), but still obtain bent tips when the conditions of significant plastic flow are met.

IV. SCANNING TUNNELING MICROSCOPE IMAGES

Surprisingly, the bent or convolved tips often produced decent scanning tunneling microscope images—of comparable quality with images produced by "pointed" tips, such as the one in Fig. 3. Subjectively, the pointed tips may have produced slightly sharper scanning tunneling microscope images, but the variation of image quality for various pointed tips was comparable with the differences between images for pointed and bent tips.

In contrast, tips that were crashed clumsily into the surface no longer produced images. (Controlled and gentle crashes, however, can leave the tips capable of forming subsequent images.[6])

V. CONCLUSIONS

Thus we conclude that the best tungsten tips are formed when only a small portion of the wire is suspended in the etchant, and that the sharpness of a tip on the ~10-μm scale of a scanning electron microscope image may not be a good indicator of tip imaging quality. Nevertheless, as a matter of good experimental practice, bent tips should be avoided, and so only a small portion of the tungsten wire should be submerged during the etching process.

ACKNOWLEDGMENTS

We are grateful for the generous support of the U.S. Army Research Office, the Defense Advanced Research Projects Agency, and the Office of Naval Research (Contracts No.

DAALO3-86G-0179, No. N0530-0716-05, and No. N00014-84-K-0352). One of us (W.J.K.) acknowledges the support of SDIO/IST.

[a] Department of Physics.
[b] Also Department of Physics; permanent address: Ford Motor Company, Dearborn, MI 48121.
[c] Present address: Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109.
[d] Department of Materials Science.
[e] College of Engineering.

[1] R. J. Colton, S. M. Baker, R. Driscoll, J. D. Baldeschwieler, and W. J. Kaiser, J. Vac. Sci. Technol., A **6**, 349 (1988) (these proceedings).
[2] R. J. Colton, S. M. Baker, J. D. Baldeschwieler, and W. J. Kaiser, Appl. Phys. Lett. **51**, 305 (1987).
[3] H. A. Mizes and W. A. Harrison, J. Vac. Sci. Technol. A **6**, 300 (1988) (these proceedings).
[4] R. M. Feenstra and A. P. Fein, Phys. Rev. B **32**, 1394. See also, R. M. Feenstra and J. A. Stroscio, J. Vac. Sci. Technol. B **5**, 923 (1987).
[5] D. U. Biegelsen, F. A. Ponce, J. C. Tramortana, and S. M. Koch, Appl. Phys. Lett. **50**, 696 (1987).
[6] R. C. Jaklevic and L. Elie, Phys. Rev. Lett. (to be published).
[7] C. J. Chen, J. Vac. Sci. Technol. A **6**, 319 (1988) (these proceedings).
[8] E. W. Müller and T. T. Tsong, *Field Ion Microscopy* (Elsevier, New York, 1969), p. 122; see also Y. Yashino, Oyo Butsuri (Appl. Phys. Jpn.) **33**, 912 (1964).

Journal of Microscopy, Vol. 152, *Pt* 3, *December* 1988, *pp.* 715–725.
Received 5 *July* 1988; *accepted (revised)* 13 *August* 1988

Nano-machining of gold and semiconductor surfaces

by WILLIAM E. PACKARD*, YOUNG LIANG*, NING DAI*, JOHN D. DOW*, RUTH NICOLAIDES†, ROBERT C. JAKLEVIC*‡ *and* WILLIAM J. KAISER*§, **Department of Physics, University of Notre Dame, IN* 46556, †*U.S. Army ARDEC, Dover, NJ* 07801, ‡*Ford Motor Company, Dearborn, MI* 48121 *and* §*Jet Propulsion Laboratory, Caltech, Pasedena, CA* 91109, *U.S.A.*

KEY WORDS. Nano-machining, diluted magnetic semiconductors, $Hg_{1-x}Cd_xTe$, $Hg_{1-x}Mn_xTe$.

SUMMARY

Using a scanning tunnelling microscope tip formed by cutting a platinum wire, we have modified the surfaces of gold and $Hg_{1-x}Cd_xTe$ on a nanometre scale by mechanical contact between the tip and the surface. By using the same tip to form images, we have been able to gain 'before' and 'after' pictures of surfaces that have been selectively 'sanded', controllably 'chiselled', and 'swept'.

We have also obtained images taken under glycerin of $Hg_{1-x}Cd_xTe$ and of the diluted magnetic semiconductor $Hg_{1-x}Mn_xTe$.

1. INTRODUCTION

Since its birth, the scanning tunnelling microscope (STM) has proven to be an exceptionally valuable tool not only for observing surfaces on the scale of atomic dimensions, but also for modifying surfaces on a nanometre scale—nano-machining. One potential application for nano-machined surfaces is in high-density information storage devices: a storage density of one bit per square nanometre would allow the text information of an entire library to be kept in a few square millimetres of area.

The STM has been used to 'machine' surfaces in several ways. Becker *et al.* (1987) wrote a bit having atomic dimensions on a Ge surface. Ringer *et al.* (1985) scribed lines on a Pd-Si surface. McCord & Pease (1986, 1987a–c) exposed resists using the STM and also used the tip as a micro-mechanical tool-bit to shave off resist material from surfaces. Abraham *et al.* (1986), Gimzewski & Möller (1987), Gimzewski *et al.* (1987) and Jaklevic & Elie (1988) used the tip of an STM to deform regions of surfaces by point contact between the tip and surface. Staufer *et al.* (1987, 1988a, b) locally melted surfaces of metallic glasses with the STM. Schneir & Hansma (1987), Schneir *et al.* (1988), Emch *et al.* (1988) and Knipping *et al.* (1988) formed hillocks and/or holes on gold surfaces by increasing the bias potential between the STM tip and the surface. Silver *et al.* (1987) and Ehrichs *et al.* (1988) used organometallic gases to write metallic features on surfaces, and de Lozanne *et al.* (1988) formed text characters on a Si substrate using

a contamination resist. Lin *et al.* (1987) etched lines on n-GaAs in an electrochemical cell. Foster *et al.* (1988) manipulated organic molecules on graphite, and Dovek *et al.* (1988) modified polymers. McCormick *et al.* (1988) produced grooves on an Al substrate using an STM.

In this paper we shall present three forms of surface modification or nano-machining which can be achieved using a cut platinum wire as the STM tip. Following a discussion of tip-preparation procedures (Section 2), we present three examples of nano-machining: selective 'sanding', and controlled 'chiselling' on gold surfaces in air (Section 3), and atomic 'sweeping' on $Hg_{1-x}Cd_xTe$ in air (Section 4). 'Sanding' smooths atomic protrusions from the surface; 'chiselling' writes a line or groove on the surface; and 'sweeping' moves material from one

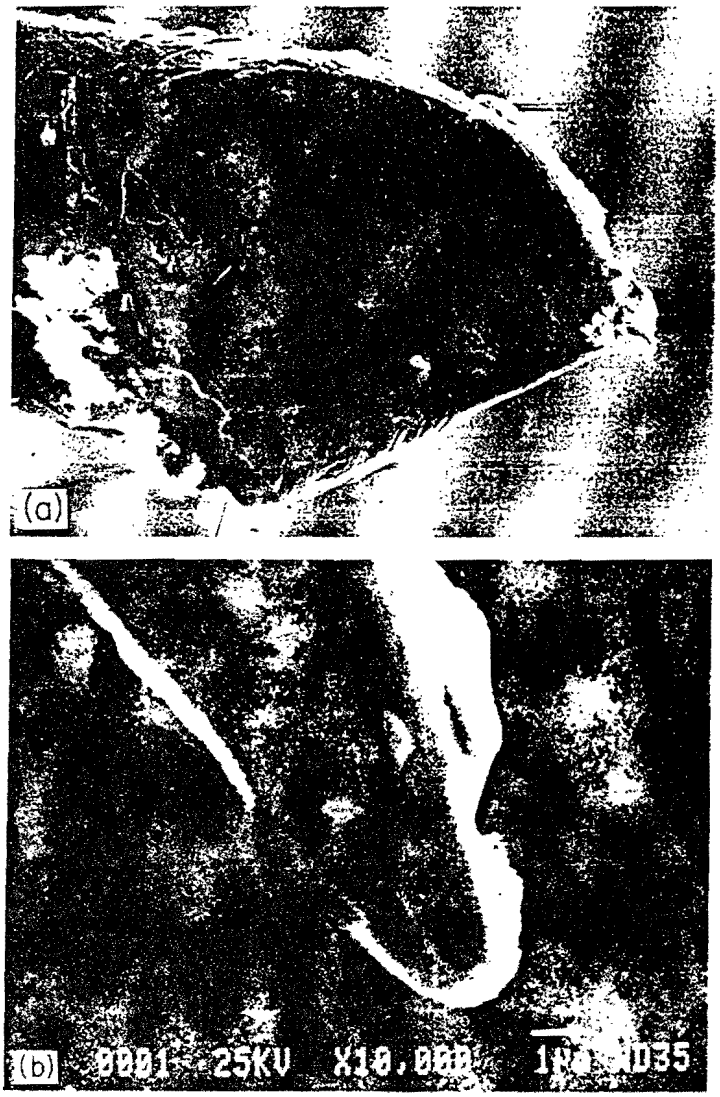


Fig. 1. SEM image of an STM tip formed by cutting a 0·25 mm Pt wire with diagonal cutter pliers. At the extreme end of the wire there is a protruding finger which is visible in both (a) low and (b) high magnification. Presumably a protruding cluster or group of atoms at the end of this finger acts as the tunnelling tip.

region to another. Section 5 presents preliminary results, including STM images of features on $Hg_{1-x}Cd_xTe$ and $Hg_{1-x}Mn_xTe$ surfaces, which were cleaved and imaged under glycerin. Finally, potential applications of nano-machining techniques will be discussed in Section 6.

2. TIP PREPARATION

The tips used in this study were prepared by cutting a 0·25 mm diameter platinum wire with diagonal cutter pliers. The angle between the long direction of the wire and the cut direction was always considerably less than 90°, so one side of the cut wire protruded more than the other side. It was hoped that there would be one predominant protrusion at the end of the wire which

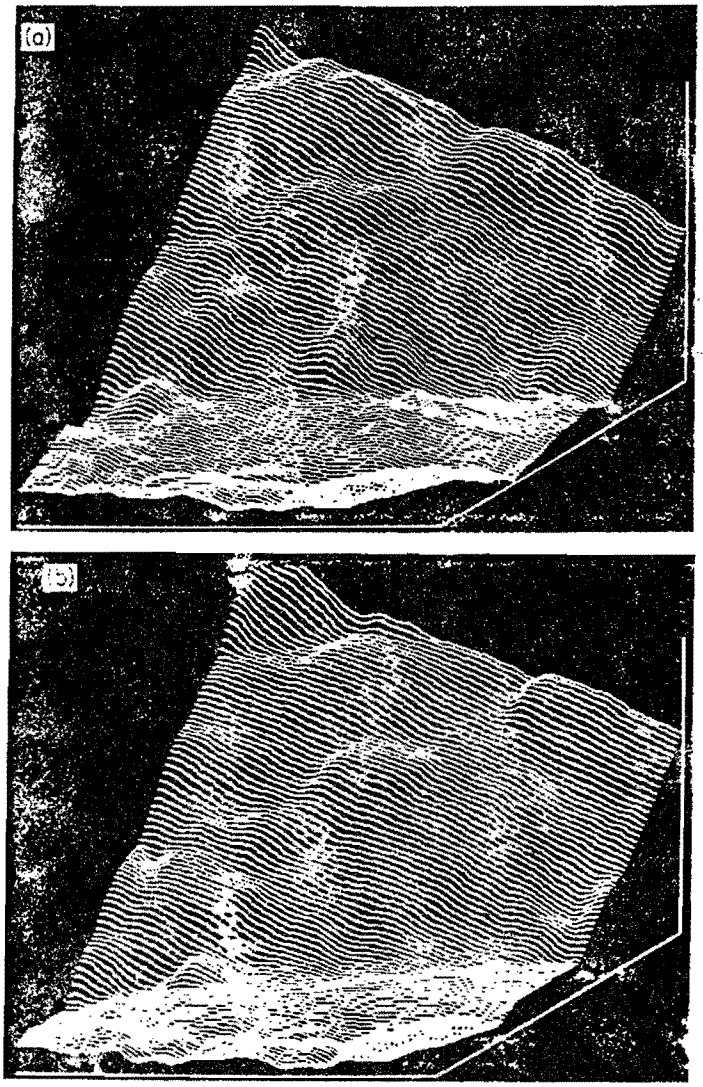


Fig. 2. STM images of a gold film (a) before and (b) after selective 'sanding'. Before 'sanding' there are two hillocks near the centre of the image. The 'sanding' operation selectively removes the two central hillocks leaving the surrounding features unchanged. The lateral range of the image is 50×50 nm (length of the horizontal lines) and the height range is 4 nm (length of the vertical line).

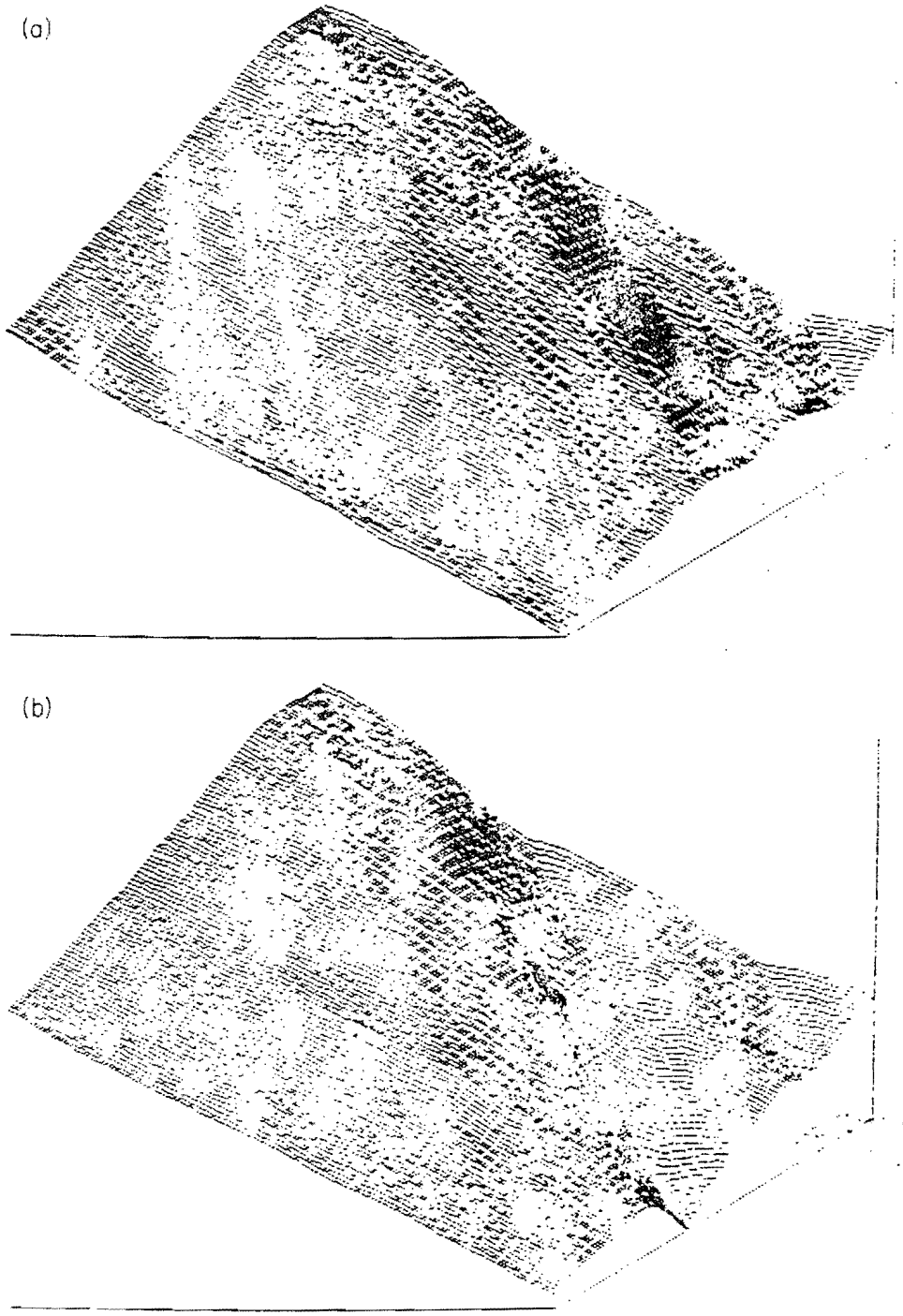


Fig. 5. STM images of a gold substrate (a) before and (b) after controlled 'chiseling'. The surface is shown after writing the (b) first, third, and (a) fourth 'chisel lines' using a Pt tip. The lateral range of the image is 200 × 200 nm (length of the horizontal lines) and the height range is 27 nm (length of the vertical lines).

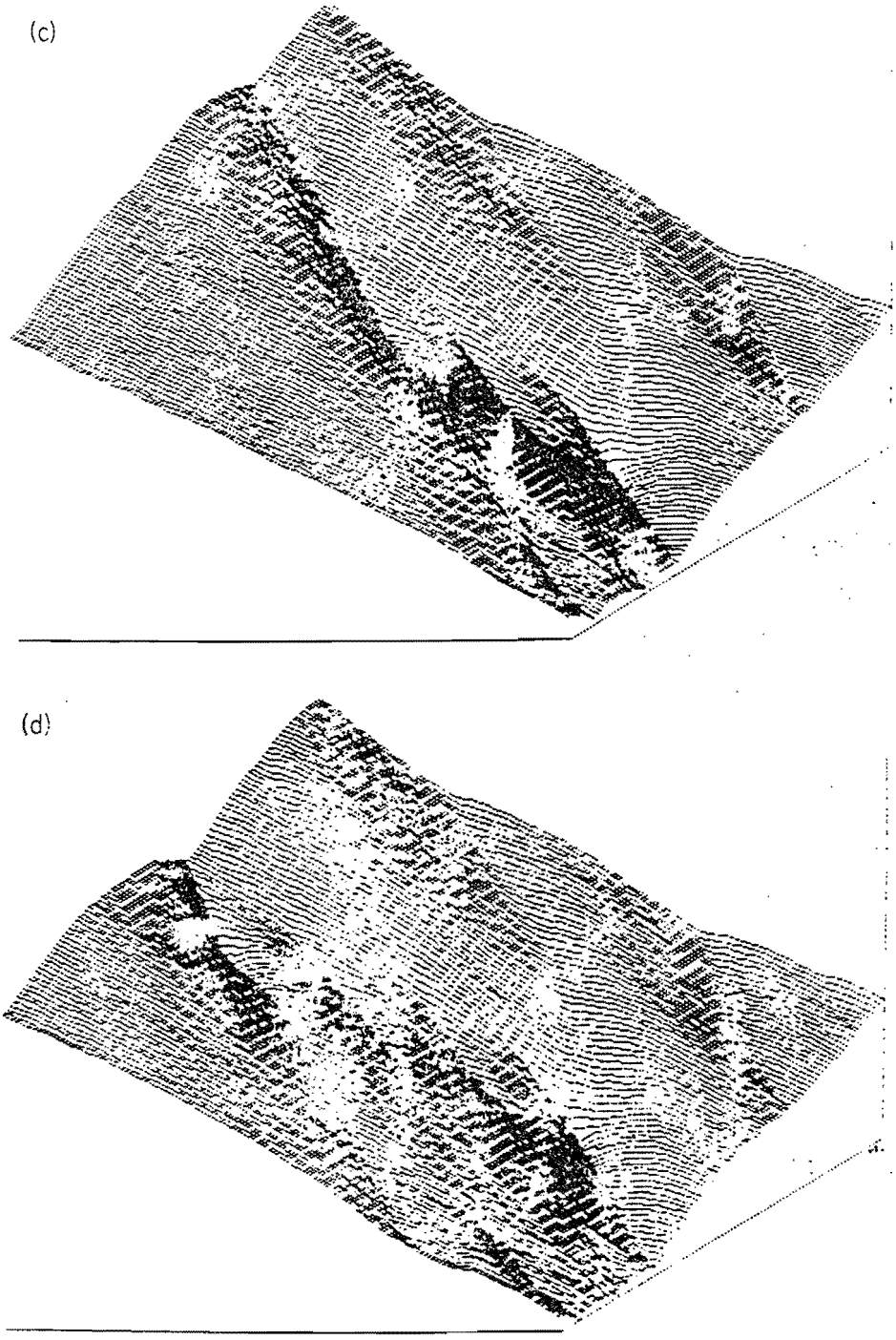


Fig. 3. continued.

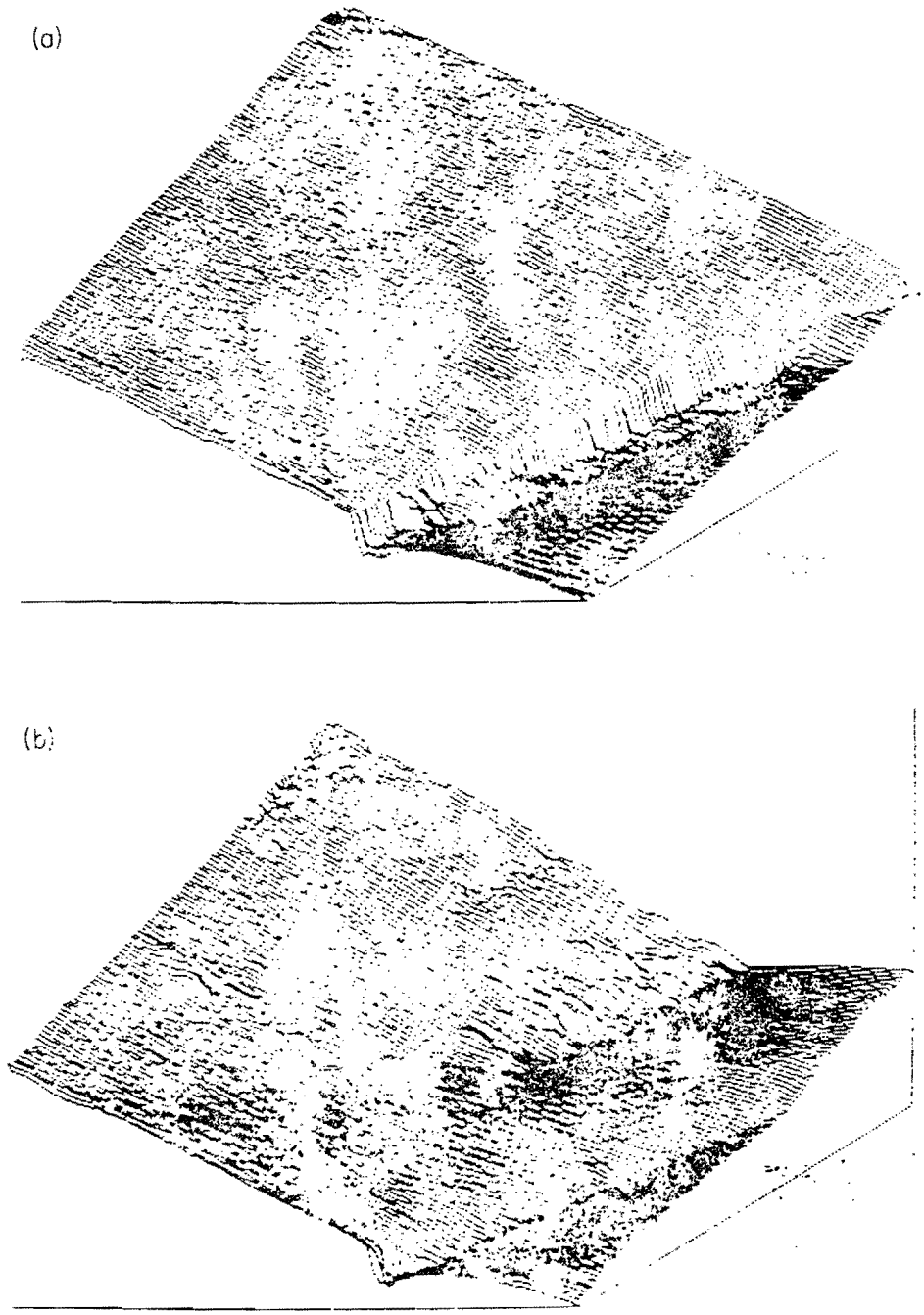


Fig. 4. STM images of a H[illegible] Te surface imaged in air (a) before tip impact, (b) immediately after the impact and (c) during the sixth post-impact scan. The tip swept material out of the crater [illegible] of tip-impact. The lateral range of the image is 30[illegible]800 nm, length of the horizontal lines, and the height range is 60 nm, length of the vertical lines.

(c)

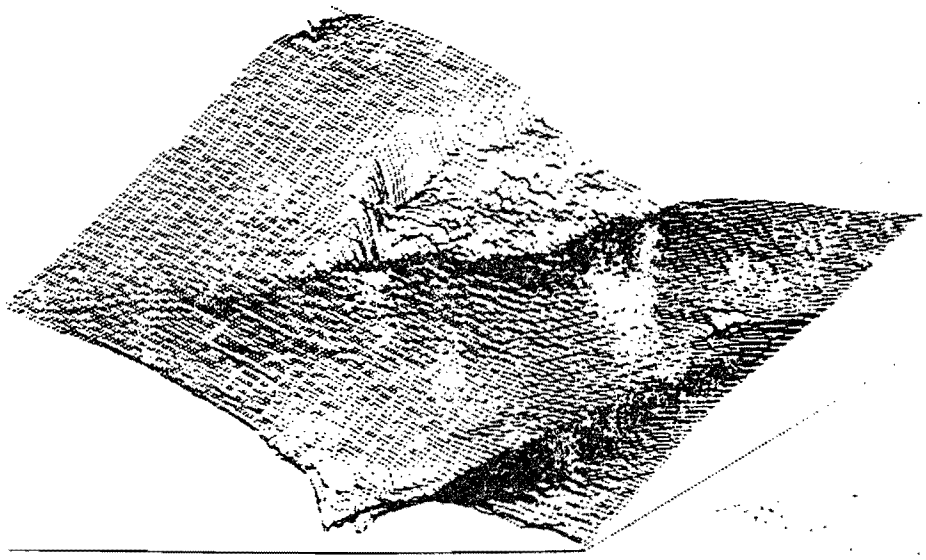


Fig. 4. continued.

would act as an STM tip and that this cut tip would be free from etchant residues normally present on etched tips. Platinum was chosen because of its relative hardness (necessary for nano-machining) and inertness to atmospheric contamination. A scanning electron microscope (SEM) image of such a cut platinum wire is shown in Fig. 1(a). (This tip had been used for controlled 'chiselling' on gold.) At the right-hand side of the figure, on the portion of the wire which was closest to the sample, a protruding finger can be seen. The finger is more clearly seen in the magnified view of Fig. 1(b), and was responsible for the current between the tip and the sample. Presumably the end of this finger is relatively flat on an atomic scale, but with a protruding cluster or group of atoms or, perhaps even a single atom, which acts as the actual STM site for electron tunnelling.

3. NANO-MACHINING OF GOLD

For these experiments we constructed an STM following the design of Kaiser & Jaklevic (1987). With this microscope we have imaged the surface of crystalline graphite, seeing well-resolved protrusions at atomic positions. Using platinum tips fabricated as described above, we have imaged several other surfaces. First we show results for Au surfaces in air.

(a) *Selective 'sanding'*

Figure 2(a) shows a 50×50 nm constant-current image taken on a gold film deposited on an insulating substrate (tip bias = + 50 mV, current = 10^{-9} A). Several small hills and features are evident. In particular, at the base of the ridge line near the horizontal centre of the picture, two hillocks can be seen with the larger one about 5 nm in diameter and about 0·3 nm high. To 'sand' these hillocks, the tip was first moved to the hillock region, then the tip was brought closer to the surface by increasing the set current a thousand-fold and finally, the hillock region was scanned at a rate too fast to allow the electronics to maintain a constant distance between the tip and the sample. Following this sanding operation, the original scanning conditions were restored and the tip was scanned over the original region of the surface. The resulting image, shown in Fig. 2(b), shows the selective removal of the two hillocks of interest while the

surrounding features remain intact, as can be seen upon comparison with Fig. 2(a). The material from the hillocks has probably been spread out over the surface by the sanding action, or perhaps some of the material may have been picked up by the sides and the end of the tip, leading to slight modification of the tip structure. Although it is not known if the hillocks were composed of gold or of some conductive contaminant, the data indicate that selective local 'sanding' (smoothing) of a surface can be accomplished with an STM tip.

(b) *Controlled 'chiselling'*

Using a platinum tip similar to the one shown in Fig. 1, we also 'chiselled' gold. A smooth gold surface was prepared by melting a piece of solid gold in a torch flame (Schneir *et al.* 1988). A constant-current scan over a 200×200 nm region of the pristine gold surface is shown in Fig. 3(a) (tip bias=+50 mV, current=10^{-9} A). The surface was modified by moving the tip over to the right-hand side of the region and then, with the position of the tip in the x–y plane of the sample fixed, the tip was moved electronically into the surface along the z-direction. Upon retraction, the tip was placed back at its original position and the region was rescanned. The resulting image (Fig. 3b) shows the same region of the surface, as indicated by the same features in the upper right-hand corner as were observed on the pristine surface. In addition, Fig. 3(b) shows a chisel line due to the impact of the tip and surface. By repeating this chiselling procedure, other lines were chiselled into the surface. Fig. 3(c) shows the surface after three such chiselling events. The chisel line to the left was done last. Subsequent chiselling ploughs up material, which partially fills in prior chisel marks. By repeating this process, four lines spaced 30–40 nm apart were chiselled into the surface (Fig. 3d). It should be emphasized that in this chiselling process the tip moves only in the z-direction, indenting the surface. This indicates that the tip geometry is probably a nanoscopic whisker (which carries the tunnelling current) on a relatively flat blade (which indents the surface).

It seems at first amazing that the STM tip can both drastically alter the topography of a region of the surface and then subsequently scan the same area. A possible explanation for this follows. After the first impact between the tip and the surface, there is undoubtably some gold which adheres to the end of the platinum wire, perhaps in the form of a little ball or 'drop'. (Energy dispersive analysis with X-rays (EDAX) of a tip (the tip of Fig. 1 which had been used for chiselling indicates the adhesion of Au to the end of the tip.) A protruding cluster of atoms on the end of this ball presumably acts as the new STM tip and is responsible for subsequent images. During chiselling, the gold on the end of the tip deforms upon contact with the gold substrate because gold is softer than platinum. This deformation continues as the platinum of the tip indents the surface. Upon retraction of the platinum tip, some gold again is picked up from the surface and a new gold drop forms at the end to the platinum thus reforming the scanning tip. The portion of the tip used for tunnelling is, therefore, geometrically different from the part creating the chisel line. The chisel line is then the mark of the platinum tip, while the electrical current passes through the softer gold tip. If the end of the platinum tip is not symmetric, then the mark of the tip is asymmetric. The data of Fig. 3 indicate an asymmetric tip, which is not surprising for our tip-fabrication method, and the line nature of the mark is very likely a result of an asymmetric platinum protrusion.

4. ATOMIC 'SWEEPING' OF $Hg_{1-x}Cd_xTe$ SURFACES

A constant-current image of a $Hg_{1-x}Cd_xTe$ ($x=0{\cdot}2$) surface in air is shown in Fig. 4(a) (tip bias=+200 mV, current=10^{-9} A). The scan was taken with the platinum tip which had previously been used to do the chiselling of Fig. 3. After the scan of Fig. 4(a), the tip impacted the surface near the centre of the picture. Immediately after impact the STM produced the image of Fig. 4(b). Although the gross features of the image, such as the ravine in the right-hand corner and the general slope of the image, remained unchanged, the details have been altered by impact, most notably, the creation of a crater. Repeated scanning of the impact area caused the material in the crater to disappear. This crater, during the sixth post-impact scan, is shown in

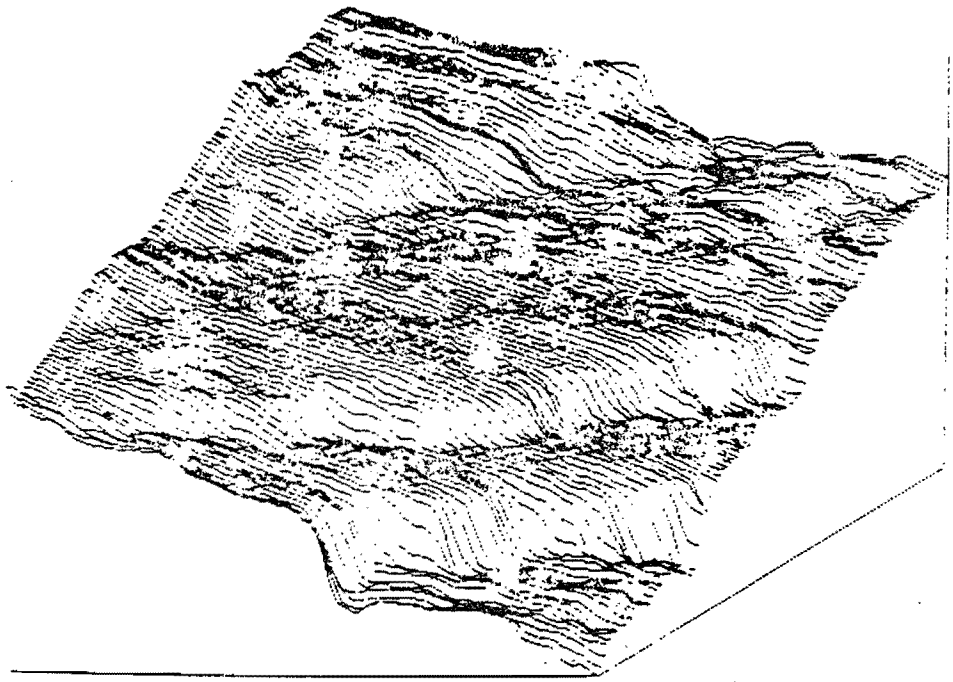


Fig. 5. STM image [illegible] $Hg_{1-x}Cd_xTe$ cleaved and imaged under glycerin. The steps are 0·5 nm and 0·8 nm high. The lateral range of the image is 20×20 nm (length of the horizontal line), and the height range is 3·5 nm (length of the vertical line).

Fig. 4(c). The emptying of the crater cannot be explained by changes (such as sharpening) of a blunted tip, since then other features such as the ravine of Fig. 4(c) would have also become more pronounced, which did not happen (compare Figs. 4b and c). Although the mechanism responsible for removal of material from the crater is unknown, a possible explanation is that the impact of the tip broke up material without dislodging it. Subsequent scanning then swept the material out of the crater considerably by either mechanical contact between the tip and the material or by electric-field induced forces on the material.

5. II–VI COMPOUNDS OBSERVED UNDER GLYCERIN

We have imaged $Hg_{1-x}Cd_xTe$ ($x=0·2$) and the dilute magnetic semiconductor $Hg_{1-x}Mn_xTe$ ($x=0·025$) after cleaving under glycerin. We hope that the glycerin (which is essentially non-conducting at the bias voltage used) acts as a protective layer on the surface, helping to prevent atmospheric contamination. A constant current scan of polycrystalline $Hg_{1-x}Cd_xTe$ cleaved and imaged under glycerin is shown in Fig. 5 (tip bias = +200 mV, current = 10^{-9} A). In the lower right [illegible] are three steps of height 0·5 and 0·8 nm. Since the macroscopic crystal was polycrystalline, the face of the $Hg_{1-x}Cd_xTe$ grain being viewed is unknown; moreover, the $Hg_{1-x}Cd_xTe$ surface could be contaminated. We cannot, therefore, state how many atomic layers the steps represent or if the steps are intrinsic $Hg_{1-x}Cd_xTe$ or a contaminant. Nevertheless, we can say that the steps would have a height of three and five atomic layers, respectively, if the exposed face were 100 [illegible] Cd_xTe.

An example of features seen on the dilute magnetic semiconductor (DMS) material $Hg_{1-x}Mn_xTe$ ($x=0·025$) are shown in Fig. 6 (tip bias = +200 mV, current = 10^{-9} A) where the sample was cleaved and imaged under glycerin. Ridges several nanometres high can be seen on the surface.

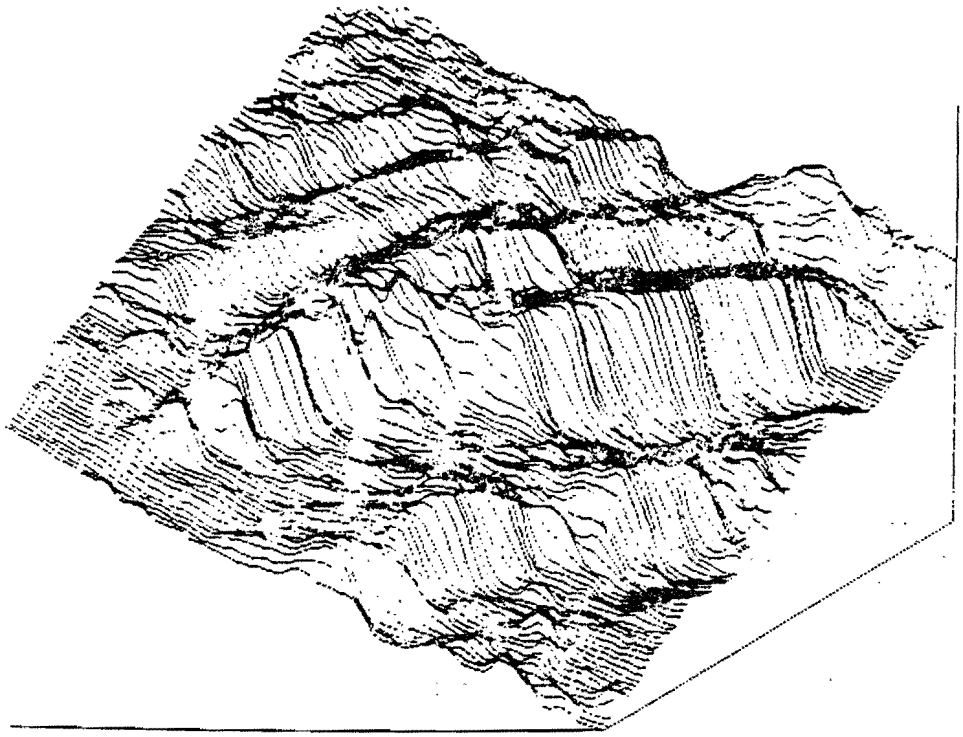


Fig. 6. Ridges on the DMS material $Hg_{1-x}Mn_xTe$ cleaved and imaged under glycerin. The ridges are several nanometres high. The lateral range of the image is 50 nm (length of the horizontal line), and the height range is 15 nm (length of the vertical line).

To our knowledge these are the first STM images of $Hg_{1-x}Cd_xTe$ or of a diluted magnetic semiconductor. To learn about the surface structures of these materials, it is clearly desirable to perform experiments in ultra-high vacuum (UHV) in order to prevent contamination. Work is presently underway in our laboratory to observe cleaved single crystals of II–VI compounds under UHV conditions. The above examples indicate that surface morphology of II–VI materials can indeed be imaged under glycerin, and strongly suggest that UHV imaging of these materials is likely to be successful also.

6. CONCLUSION

We have shown three examples of surface alterations induced by STM tips consisting of cut platinum wire: selective 'sanding' of Au, controlled 'chiselling' of Au, and atomic 'sweeping' of $Hg_{1-x}Cd_xTe$ in air. It is possible to smooth selected regions of the surface by using the sanding action of the STM tip, while leaving neighbouring structures essentially unchanged. Controlled 'chiselling' of lines 30–40 nm apart has been achieved on a gold surface using a cut platinum wire. It was shown that the STM tip can 'sweep' or move material out of a tip-impact area on the surface of $Hg_{1-x}Cd_xTe$.

In summary, these machining operations, although primitive, could be used in refined forms for fabrication of information storage structures. For example, the chiselling operation performed with a more rounded tip would produce more circular craters on gold. The tip could then be used under computer control to perform dot matrix printing on the surface, producing a written pattern, or perhaps even a picture, which could be erased by using the selective 'sanding' operation. The sweeping and sanding operations could possibly be used to transport groups of atoms from one region of the surface to another, 'herding' atoms or 'painting' a line on.

a surface. In addition we have presented, to our knowledge, the first STM images of $Hg_{1-x}Cd_xTe$ and a DMS material.

ACKNOWLEDGMENTS

We thank J. K. Furdyna for DMS samples. We are grateful for the generous support of the U.S. Army Research Office and the Defense Advanced Research Projects Agency (Contract Nos. DAAL03-87-K-0112 and NO530-0716-05).

REFERENCES

Abraham, D.W., Mamin, H.J. & Clarke, J. (1986) Surface modification with the scanning tunneling microscope. *IBM J. Res. Dev.* **30**, 492–498.

Becker, R.S., Golovchenko, J.A. & Swartzentruber, B.S. (1987) Atomic-scale surface modifications using a tunnelling microscope. *Nature*, 325, 419–421.

de Lozanne, A.L., Ng, K.W., Pan, S., Silver, R.M. & Berezin, A. (1988) Tunnelling spectroscopy of high-temperature superconductors. *J. Microsc.* 152, 117–122.

Dovek, M.M., Albrecht, T.R., Kuan, W.J., Lang, C.A., Emch, R., Grütter, P., Frank, C.W., Pease, R.F.W. & Quate, C.F. (1988) Observation and manipulation of polymers by scanning tunnelling and atomic force microscopy. *J. Microsc.* 152, 229–236.

Ehrichs, E.E., Silver, R.M. & de Lozanne, A.L. (1988) Direct writing with the scanning tunneling microscope. *J. Vac. Sci. Technol.* **A6**, 540–543.

Emch, R., Nogami, J., Dovek, M.M., Lang, C.A. & Quate, C.F. (1988) Characterization and local modification of anatomically flat gold surfaces by STM. *J. Microsc.* 152, 129–136.

Foster, J.S., Frommer, J.E. & Arnett, P.C. (1988) Molecular manipulation using a tunnelling microscope. *Nature*, 331, 324–326.

Gimzewski, J.K. & Möller, R. (1987) Transition from the tunneling regime to point contact studied using scanning tunneling microscopy. *Phys. Rev. B*, 36, 1284–1287.

Gimzewski, J.K., Möller, R., Pohl, D.W. & Schlittler, R.R. (1987) Transition from tunneling to point contact investigated by scanning tunneling microscopy and spectroscopy. *Surf. Sci.* **189/190**, 15–23.

Jaklevic, R.C. & Elie, L. (1988) Scanning-tunneling-microscope observation of surface diffusion on an atomic scale: Au on Au(111). *Phys. Rev. Lett.* 60, 120–123.

Kaiser, W.J. & Jaklevic, R.C. (1987) Scanning tunneling microscopy of metals: spectroscopy and topography. *Surf. Sci.* 181, 55–68.

Nagahara, L., Lindsay, S.M., Thundat, T. & Knipping, U. Tip-bias induced surface modification on gold surfaces. *J. Microsc.* 152, 145–148.

Lin, C.W., Fan, F.R. & Bard, A.J. (1987) High resolution photoelectrochemical etching of n-GaAs with the scanning electrochemical and tunneling microscope. *Electrochem. Soc.* 134, 1038–1039.

McCord, M.A. & Pease, R.F.W. (1986) Lithography with the scanning tunneling microscope. *J. Vac. Sci. Technol.* B4, 86–88.

McCord, M.A. & Pease, R.F.W. (1987a) The effect of reflected and secondary electrons on lithography with the scanning tunneling microscope. *Surf. Sci.* 181, 278–284.

McCord, M.A. & Pease, R.F.W. (1987b) Exposure of calcium fluoride resist with the scanning tunneling microscope. *J. Vac. Sci. Technol.* B5, 430–433.

McCord, M.A. & Pease, R.F.W. (1987c) Scanning tunneling microscope as a micromechanical tool. *Appl. Phys. Lett.* 50, 569–570.

McCormick, L.D., Thundat, T., Nagahara, L. & Lindsay, S.M. (1988) Abstract. In: *Proc. Royal Microscopical Society*, Vol. 23, p. 538.

Ringer, M., Hidber, H.R., Schlögl, R., Oelhafen, P. & Güntherodt, H.-J. (1985) Nanometer lithography with the scanning tunneling microscope. *Appl. Phys. Lett.* 46, 832–834.

Schneir, J. & Hansma, P.K. (1987) Scanning tunneling microscopy and lithography of solid surfaces covered with nonpolar liquids. *Langmuir*, 3, 1025–1027.

Schneir, J., Sonnenfeld, R., Marti, O., Hansma, P.K., Demuth, J.E. & Hamers, R.J. (1988) Tunneling microscopy, lithography, and surface diffusion on an easily prepared, atomically flat gold surface. *J. appl. Phys.* 63, 717–721.

Silver, R.M., Ehrichs, E.E. & de Lozanne (1987) Direct writing of submicron metallic features with a scanning tunneling microscope. *Appl. Phys. Lett.* 51, 247–249.

Staufer, U., Wiesendanger, R., Eng, L., Rosenthaler, L., Hidber, H.-R. Güntherodt, H.-J. & Garcia, N. (1987) Nanometer scale structure fabrication with the scanning tunneling microscope. *Appl. Phys. Lett.* 51, 244–246.

Staufer, U., Wiesendanger, R., Eng, L., Rosenthaler, L., Hidber, H.-R., Güntherodte, H.-J. & Garcia, N. (1988a) Surface modification in the nanometer range by the scanning tunneling microscope. *J. Vac. Sci. Technol.* A6, 537–539.

Staufer, U., Scandella, L., Anselmetti, D., Eng, L., Hidber, H.R., Jung, T., Lapka, R., Wiesendanger, R. & Güntherodt, H.-J. (1988b) Abstract. In: *Proc. Royal Microscopical Society*, Vol. 23, part 3, S28.

/mnt/dow/manus/dots266.rno March 10, 1992

Fabrication of Quantum Dots on the InSb(110) Surface

Yong Liang, William E. Packard, and John D. Dow

Department of Physics, University of Notre Dame,

Notre Dame, IN 46556

Abstract

Using a scanning tunneling microscope, quantum dots of diameter 30 Å to 50 Å have been fabricated on the cleaved InSb(110) surface in UHV. Both In and Sb atoms were clearly resolved on the cleaved surface.

PACS Index Numbers:

Scanning tunneling microscopes (STM's) present the opportunity not only to image a semiconductor surface with unprecedented resolution but also to controllably alter the surface on the Ångstrom scale, as a result of the tip-surface interaction. Becker et al. demonstrated that single atoms could be manipulated on the Ge(111) surface with a single voltage pulse [1]. A number of authors [2] have "machined" various surfaces on the nanometer scale, either by crashing the STM tip into the surface, by applying voltage pulses, or by local heating -- with experiments performed in air, under liquids, or in ultra-high vacuum. In this paper we present some of the first [3] STM images of the InSb(110) surface under ultra-high vacuum and we show how quantum dots can be controllably engraved in the surface.

We employed a Pachyderm-4 sculpted STM [4], which is machined almost in its entirety from a single block of stainless steel and hence has unusual vibrational stability. With this STM, we obtained the 160 Å × 180 Å image of Fig. 1 by scanning the InSb(110) surface. This surface was cleaved in 10^{-11} Torr vacuum, and the image was taken with a tunneling current of 120 pA under positive sample bias of +0.2 V (sensitive to electronically unoccupied states of the InSb). The white features of the image are In atoms spaced by the accepted lattice constants (assuming a 29^{o} Rigid Rotation Model [5] of the surface relaxation). The number of visible defects on this picture is one (in the upper right corner) per $\simeq 10^{3}$ atoms.

Negative sample bias of -0.2 V produced the image of Fig. 2, which is sensitive to electronically occupied Sb-derived states.

By positioning the STM tip over a spot for $\simeq 10^{2}$s we formed quantum dots, such as those of Fig. 3. The positions of the dots were easily controllable. The perfect InSb(110) surface was more susceptible to disruption by the STM

tip than other surfaces we have studied [2][6], and so preservation of the perfect surface required rapid scanning, while the formation of quantum dots was almost effortless. The diameters of the two dots in Fig. 3 are ≃ 30 Å and ≃ 50 Å, namely, 5 to 8 times the InSb lattice constant. Close inspection of that figure reveals images of individual In atoms.

The quantum dots appear to be small surface voids, with depths of 2 Å to 4 Å. The size of the voids can be increased by holding the tip fixed for a longer time or by scanning over a very small area.

These results demonstrate that Ångstrom-scale STM lithography of InSb(110) surfaces will be feasible, and that it might be possible, if the electronic structures of the quantum dots are favorable, to fabricate Ångstrom-scale memory bits on this surface.

Acknowledgment -- We are grateful to the U.S. Army Research Office for their support of this work (Contract No. DAAL03-87-K-0112). We have also benefited from stimulating discussions of STM design with R. Jaklevic, W. Kaiser, and S.-L. Tang, and we thank B. Swartzentruber for a copy of his STM software.

REFERENCES

[1] R. S. Becker, J. A. Golovchenko, B. S. Swartzentruber, Nature 325, 419 (1987).

[2] W. E. Packard, Y. Liang, N. Dai, J. D. Dow, R. Nicolaides, R. C. Jaklevic, and W. J. Kaiser, J. Microscopy 152, 715 (1988) and references therein.

[3] See Y. Liang, W.-M. Hu, W. E. Packard, and J. D. Dow, Bull. Am. Phys. Soc. 35, 227 (1990). L. J. Whitman, J. A. Stroscio, R. A. Dragoset, Bull. Am. Phys. Soc. 35, 226 (1990) have independently reported images of the InSb(110) surface recently.

[4] Pachyderm Industries technical report TR1000-01.

[5] S. Y. Tong, A. R. Lubinsky, B. J. Mrstik, and M. A. Van Hove, Phys. Rev. B 17, 3303 (1978); D. J. Chadi, Phys. Rev. B 18, 1800 (1978); 19, 2074 (1979); D. V. Froelich, M. E. Lapeyre, J. D. Dow, and R. E. Allen, Superlatt. Microstruct. 1, 87-89 (1985); R. V. Kasowski, M.-H. Tsai, and J. D. Dow, J. Vac. Sci. Technol. B 5, 953-5 (1987).

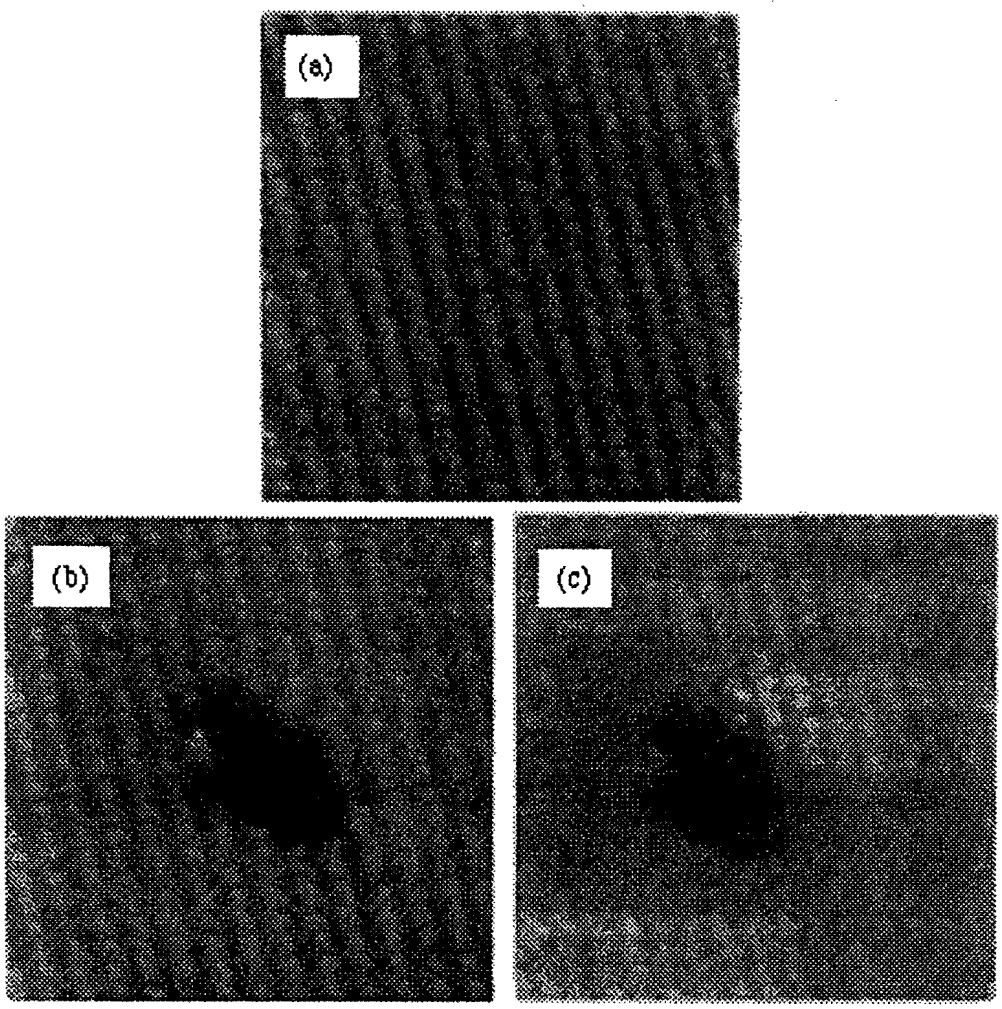
[6] W. E. Packard, N. Dai, and J. D. Dow, J. Vac. Sci. Technol. B, to be published.

FIGURE CAPTIONS

Fig. 1. STM image of a 160 Å × 180 Å area of InSb(110). The surface was cleaved in ultra-high vacuum (10^{-11} torr range) and imaged at +0.2 V sample bias with 120 pA tunneling current. The white features are In atoms. The data were Fourier filtered.

Fig. 2. STM image of InSb(110) showing Sb atoms. The sample bias was -0.2 V and the tunnel current was 120 pA. Individual Sb atoms are clearly resolved. The data were Fourier filtered.

Fig. 3. Quantum dots produced on the InSb(110) surface. Two quantum dots were produced on a perfect InSb surface by placing the tip over each dot region for $\simeq 10^2$ seconds. The dots are $\simeq$ 30 Å and $\simeq$ 50 Å in diameter. Individual In atoms can also be observed in the image. The sample was biased +0.5 V and the tunneling current was 120 pA. The data were not filtered.



STM image showing the surface topography before (a) and after (b) (c) the nano-fabrication. The sample bias is 0.3 V (a) (b) and -0.3 V (c).

Surface Science 228 (1990) 49–52
North-Holland

DEEP LEVELS AND SHALLOW–DEEP TRANSITIONS IN ZnSe/$Zn_{1-x}Mn_xSe$ SUPERLATTICES

Shang Yuan REN, Jun SHEN, Run-Di HONG, Stefan KLEMM, M.-H. TSAI and John D. DOW
Department of Physics, University of Notre Dame, Notre Dame, IN 46556, USA

Received 2 June 1989; accepted for publication 18 July 1989

A theory of deep impurity levels in superlattices is outlined, and applied to ZnSe/$Zn_{0.5}Mn_{0.5}Se$ with the substitutional dopant Ga_{Zn}. Ga is predicted to undergo a shallow–deep transition as a function of x in bulk $Zn_{1-x}Mn_xSe$, and so Ga dopes ZnSe n-type but prevents $Zn_{0.5}Mn_{0.5}Se$ from being doped n-type even by modulation doping. In ZnSe/$Zn_{0.5}Mn_{0.5}Se$ superlattices, the band edges are quite sensitive to changes in the layer thicknesses, but the deep levels are not. As a result, shallow–deep transitions as functions of layer thickness are predicted to occur. The physics of shallow–deep transitions in superlattices is elucidated, and its relevance to the II–VI doping problem is discussed.

1. Introduction

Every s- and p-bonded substitutional impurity in a semiconductor produces four "deep" levels that lie near or in the fundamental band gap of the host. These levels are due to the central-cell defect potential, and may all lie resonant with the host bands, in which case the impurity is termed "shallow." Or at least one of these levels may lie within the gap, in which case the impurity is "deep" [1]. Normally one such deep level is A_1-symmetric or s-like and three are p-like (and possibly degenerate, depending on the site symmetry).

2. Shallow–deep transitions

In a very crude (but instructive) approximation [2], the deep levels are insensitive to changes of the host composition, atomic ordering (e.g., superlattice versus random alloy), or pressure, and retain their absolute energies. In contrast, the conduction and valence band edges are sensitive to such changes, and so it is rather common that a band edge passes through a deep level, changing the character of the impurity from shallow to deep. This is believed to be the case [3] for Si_{cation} (Si on a cation site) in $Al_xGa_{1-x}As$: for $x < 0.2$ the A_1-symmetric deep level lies in the conduction band, making Si a shallow donor; but for $x > 0.3$ the deep level is in the fundamental band gap, allowing the Si atom to trap an extra electron rather than donate one to the conduction band, rendering the material semi-insulating rather than n-type [3]. This shallow–deep transition is particularly interesting in superlattices, where the band edges are sensistive to the choice of layer thickness, but the deep levels are not [2].

3. Shallow–deep transitions in superlattices

One example of such a transition is the Ga_{Zn} impurity near the center of a ZnSe layer in a ZnSe/$Zn_{0.5}Mn_{0.5}Se$ [001] superlattice. This impurity is a shallow donor, with its A_1-symmetric deep level in the conduction band, for thick ZnSe layers. Fig. 1 illustrates how the conduction band edge of a $N \times 10$ superlattice passes through the Ga deep level as N decreases from $N = 10$ to $N = 1$ (a single layer of ZnSe). In the thin superlattices (for $N < 3$, according to the theory [2,4]), the superlattice's conduction band edge lies above the deep level; and Ga becomes a deep impurity: the extra electron (relative to Zn) of neutral Ga is trapped in the deep level, which can also trap an additional electron of opposite spin. For thick ZnSe layers ($N \geq 3$), the extra electron of neutral

0039-6028/90/$03.50

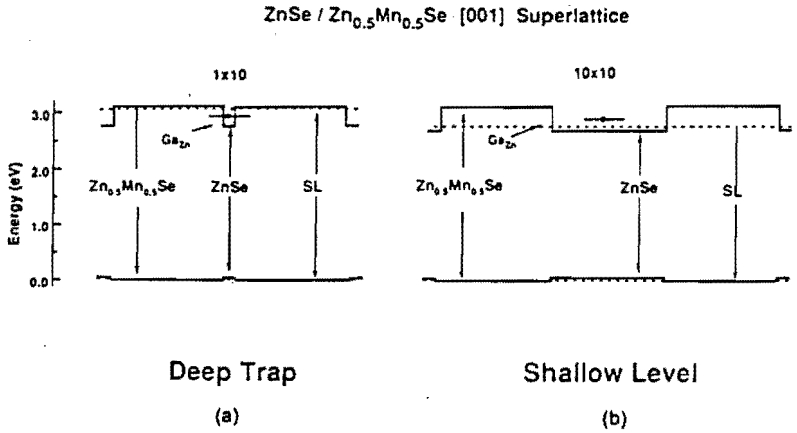


Fig. 1. Well-center Ga_{Zn} A_1 deep levels in (a) thin (1×10) and in (b) thick (10×10) quantum well ZnSe/$Zn_{0.5}Mn_{0.5}Se$ [001] superlattices. The thick solid lines are the band edges of ZnSe and $Zn_{0.5}Mn_{0.5}Se$ bulk semiconductors respectively. The dashed lines are the superlattice band edges. The short solid lines are the Ga_{Zn} deep levels. The Ga_{Zn} deep level is above the superlattice conduction band edge in a 10×10 superlattice and is in the gap of a 1×10 superlattice. The extra valence electron in the Ga_{Zn} resonant deep level will fall to the conduction band edge in the 10×10 superlattice, while in the 1×10 superlattice the extra electron will occupy the deep level which can also trap another electron of opposite spin.

Ga spills out of the deep level (which lies above the superlattice's conduction band edge) and the Ga is autoionized, creating a long-ranged Coulomb potential which binds the electron at zero temperature in a shallow donor level.

Fig. 2 illustrates how the Ga deep level, the conduction band minimum (CBM), the valence band maximum (VBM), and the shallow level are predicted to vary with ZnSe layer thickness N in an $N \times 10$ ZnSe/$Zn_{0.5}Mn_{0.5}Se$ superlattice. The predictions use an empirical tight-binding Hamiltonian [5,6] together with the Green's function method [1].

This behavior of the Ga deep level as a function of layer thickness N is similar to that found as a function of alloy composition x in $Zn_{1-x}Mn_xSe$: for $x > 0.1$ the Ga deep level lies in the band gap, not in the conduction band and traps electrons rather than donating them. This means that doping of $Zn_{1-x}Mn_xSe$ for $x > 0.1$ with Ga should produce semi-insulating rather than n-type material, which appears to be the case experimentally [7]. Even modulation doping of $Zn_{1-x}Mn_xSe$ with Ga will not produce n-type material for $x > 0.1$, because Ga is a deep trap in both layers of a $Zn_{1-x}Mn_xSe/Zn_{1-y}Mn_ySe$ superlattice for $y > x > 0.1$.

Fig. 3 illustrates the predicted dependence on alloy composition x of the levels of a Ga_{Zn} impurity in the ZnSe layer of a 1×10 ZnSe/$Zn_{1-x}Mn_xSe$ superlattice. For $x = 0$, the superlattice reduces to bulk ZnSe, and Ga has a shallow hydrogenic ground state donor level slightly below the conduction band minimum, which provides n-type doping. The Ga deep level lies above the conduction band minimum. As the alloy composition x of the ZnSe/$Zn_{1-x}Mn_xSe$ superlattice increases, the band gap opens up and the conduction band edge (measured with respect to the valence band maximum) moves to higher energy until, near $x = 0.4$, the band edge passes through the deep level. For $x > 0.4$, the stable ground state of the neutral Ga impurity in the 1×10 ZnSe/$Zn_{1-x}Mn_xSe$ superlattice has the deep level occupied by one electron. This deep level can trap a second electron of opposite spin, and so it removes electrons from the conduction band, making the material semi-insulating rather than n-type.

Shallow–deep transitions can occur when the valence band edge passes through a deep level, much the same as when a conduction band edge does. The valence-band shallow–deep transitions normally have a much more dramatic effect on the

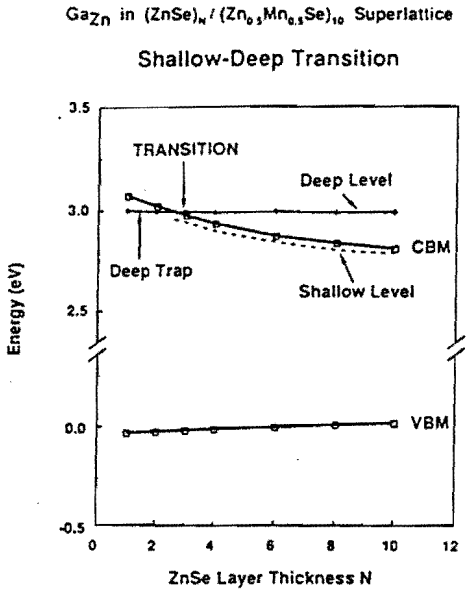


Fig. 2. Dependence of deep levels and superlattice band edges on ZnSe layer thickness N (number of ZnSe molecules thick) for $N \times 10$ $ZnSe/Zn_{0.5}Mn_{0.5}Se$ superlattices. The thick lines are the superlattice conduction band (CBM) and valence band (VBM) edges respectively. The top of the ZnSe valence band (without strain) is taken to be the zero of energy. The thin solid line is the predicted well-center Ga_{Zn} A_1 deep level. The shallow level, which follows the conduction band edge, is depicted by a dashed line. A shallow–deep transition is predicted to occur around $N \approx 3$.

doping character of a material, however, because they invariably involve p-like deep levels capable of containing six electrons (whereas the conduction-band shallow–deep transitions normally involve A_1-symmetric levels capable of trapping only two electrons).

We believe that such shallow–deep transitions are responsible for the different doping characters [8,9] of ZnSe (which can be easily doped n-type but not p-type) and ZnTe (which can be doped p-type): deep levels that lie in the gap of ZnSe and trap holes instead lie below the valence band maximum in ZnTe and donate holes. Clearly one way to enhance the p-type dopability of a II–VI semiconductor is to manipulate the semiconductor's valence band maximum, moving it up in energy until it covers the deep hole traps. For example, the p-dopability of CdTe can be improved [10] by fabricating a CdTe/ZnTe strained-layer superlattice. In this case the strain splits the

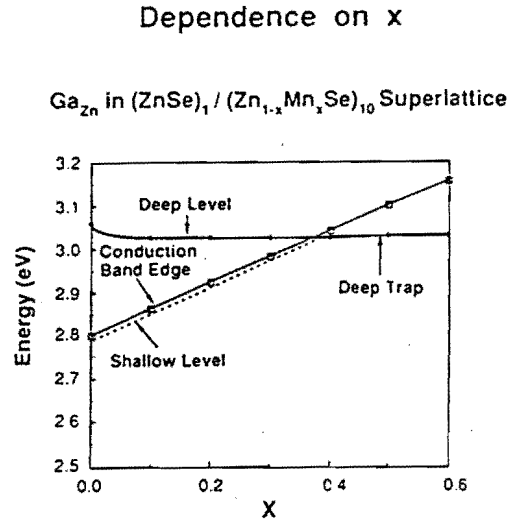


Fig. 3. Predicted dependence on Mn concentration x of the Ga_{Zn} deep level, and the shallow donor level in 1×10 $ZnSe/Zn_{1-x}Mn_xSe$ [001] superlattices. The Ga_{Zn} A_1 deep level is resonant with the conduction band when $x < 0.4$ (making Ga a shallow donor impurity), and is a deep trap occupied by the extra electron for $x > 0.4$.

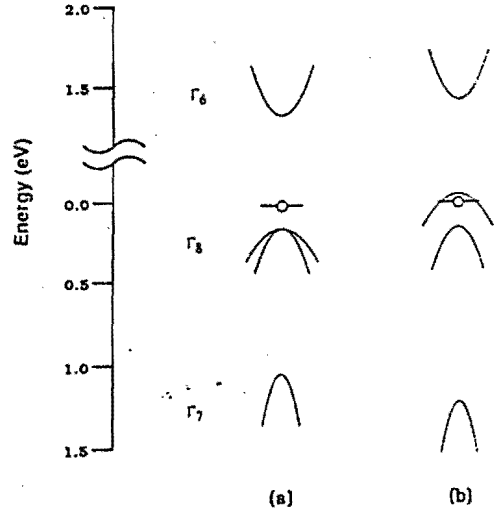


Fig. 4. Schematic energy band structure (energy in eV versus wave vector) of CdTe, illustrating how strain qualitatively changes the valence band level structure with respect to the deep level energy and covers the deep level. (a) The bulk semiconductor with a deep hole trap (that also contains at least one hole) within 0.2 eV of the valence band edge. The Γ_7 and Γ_8 bands are p-like bands that are split due to the large spin–orbit interaction in CdTe. (b) A 2×4 superlattice has an internal strain that further splits the valence band and covers up the deep level, autoionizing the hole.

valence band maximum of the CdTe and covers up deep hole traps in the gap slightly above the valence band maximum (fig. 4). A more complete discussion of this p-doping problem will be published elsewhere [11].

4. Summary

The physics of shallow–deep transitions plays a major role in determining the doping properties of II–VI semiconductors. Band edges pass through deep levels and change the doping character of the impurity from n-type (donor) to semi-insulating (trap) or from p-type (acceptor) to semi-insulating. We believe that by better understanding and using these shallow–deep transitions, it will be possible to circumvent many of the doping problems that currently plague II–VI semiconductors.

Acknowledgments

We are grateful to the Office of Naval Research, the Air Force Office of Scientific Research, the Army Research Office, and the Defense Advanced Research Projects Agency for their generous support (Contract Nos. N00014-89-J-1136, Af-AFOSR-89-0063, DAAL03-87-K-0112, and N0530-0716-05).

References

[1] H.P. Hjalmarson, P. Vogl, D.J. Wolford and J.D. Dow, Phys. Rev. Lett. 44 (1980) 810;
for the concepts that form the foundation of this work, see also W.Y. Hsu, J.D. Dow, D.J. Wolford and B.G. Streetman, Phys. Rev. B 14 (1977) 1597.
[2] S.Y. Ren, J.D. Dow and J. Shen, Phys. Rev. B 38 (1988) 10677.
[3] H.P. Hjalmarson and T.J. Drummond, Appl. Phys. Lett. 48 (1986) 657; see also, ref. [1].
[4] J. Shen, S.Y. Ren and J.D. Dow, to be published.
[5] P. Vogl, H.P. Hjalmarson and J.D. Dow, J. Phys. Chem. Solids 44 (1983) 365.
[6] A. Kobayashi, O.F. Sankey and J.D. Dow, Phys. Rev. B 28 (1983) 946.
[7] R. Gunshor and G. Prinz, private communications.
[8] R.N. Bhargava, J. Cryst. Growth 59 (1982) 15, and references therein.
[9] See, for example, C.H. Henry, K. Nassau and J.W. Shiever, Phys. Rev. B 4 (1971) 2453, and references therein.
[10] S.Y. Ren, J.D. Dow and S. Klemm, J. Appl. Phys., in press.
[11] J.D. Dow, Run-Di Hong, S. Klemm, S.Y. Ren and O.F. Sankey, to be published.

Reprinted from: Journal of Vacuum Science & Technology B, Part II, Vol. 9, No. 2, Mar.–Apr. 1991

Scanning tunneling microscopy study of the cleaved InSb(110) surface

Yong Liang, William E. Packard, and John D. Dow
Department of Physics and Astronomy, Arizona State University, Tempe, Arizona 85287

Scanning tunneling microscopy study of the cleaved InSb(110) surface

Yong Liang, William E. Packard, and John D. Dow
Department of Physics and Astronomy, Arizona State University, Tempe, Arizona 85287

(Received 3 August 1990; accepted 18 December 1990)

The clean InSb(110) surface was imaged in ultrahigh vacuum with scanning tunneling microscopy. A 1×1 surface structure was observed. A super-periodicity consistent with a $c(4\times6)$ reconstruction was also observed on some regions of some cleaved surfaces, and appears to be cleavage dependent. The InSb(110) surface can easily be altered by the tunneling process to produce nanoscopic dots on the surface as small as 9 Å radius.

I. INTRODUCTION

Scanning tunneling microscopy (STM) presents the opportunity not only to image a semiconductor surface on the atomic scale[1] but also to controllably manipulate the surface atoms.[2] In this paper we present some of the first atom-resolved STM images of cleaved InSb(110) taken in ultrahigh vacuum (UHV).[3,4] We observed the expected 1×1 structure similar to GaAs(110).[5] In addition we also observed a new surface structure on some regions or domains of the surface. This new structure is consistent with a $c(4\times6)$ reconstruction. Furthermore, we found that the InSb(110) surface is very sensitive to disruption by the tunneling process, a fact that allowed us to controllably produce nanoscopic dots on the surface.

II. EXPERIMENTAL

For these studies we used a Pachyderm-4 UHV scanning tunneling microscope[6] which incorporates a demountable sample holder capable of holding six samples at one time. This microscope has unusual vibrational immunity because it was sculpted, almost in its entirety, from a single block of stainless steel. The microscope head was bolted directly onto the vacuum flange, and all the vibration isolation was done exterior to the vacuum chamber. Individual samples were *n*-type single crystal bars purchased from Sumitomo, which were 5 mm long with a 1×3 mm^2(110) surface cross section. Samples were cleaved and imaged in a vacuum of order $\sim10^{-11}$ Torr. The cleaved (110) surfaces were typically mirror-like and flat, with few visible macroscopic steps. For all the pictures shown in this paper the cleavage direction was from the bottom left corner of a picture to the top right corner. Since the samples were cantilevered, the cleavage knife never contacted the cleaved surface because the samples always cleaved where the samples were fastened to the sample holder and not where the knife contacted them. The W tips of the STM were electrochemically etched in NaOH solution followed by ultrasonic cleaning in deionized water. No further tip cleaning procedures were performed in vacuum. These tips, almost without fail, gave atomic-resolution images. Typical scan rates were about 500 Å/s with a current of about 100 pA.

III. RESULTS AND DISCUSSION

A. Surface structure of cleaved InSb(110)

On the InSb(110) surface we observed the expected 1×1 surface structure similar to GaAs(110). Large defect-free or nearly defect-free regions of this structure were frequently observed. A typical example of such an area (150×150 Å^2) is shown in Fig. 1, where the white features are individual In-derived states (sample biased $+0.2$ V). Sb-derived states (golden features) were observed under negative sample bias (-0.2 V) as shown in Fig. 2 (35×35 Å^2). The spacing of the observed 1×1 unit cell is in agreement with the accepted lattice constants of 6.5×4.6 Å.

In addition to the 1×1 structure we also observed a super-periodicity on some regions of the InSb(110) surface for some cleaves. This super-periodicity consisted of rows spaced 19 Å apart and inclined approximately 45° with respect to the In or Sb rows. An example of this super-periodicity is shown in Fig. 3 where two sets of rows can be observed: one vertical, closely spaced set and another diagonal, widely spaced set. The more closely spaced rows have a 6.5 Å spacing consistent with the In–Sb zig–zag chain spacing. We have taken precautions to guarantee that the super-periodicity is not an artifact or due to noise, such as 60 Hz pickup: The row spacing and relative orientation did not change with scan parameters such as the scan rate or with the scan direction. For example, changing the scan rate by a factor of more than 2 caused no change in the row spacings or their relative angles, and halfing the scan range while keeping the scan rate constant caused the row spacing in a picture to scale with the range. Furthermore, when we changed the tip scan direction by 20° to 90°, the STM pictures rotated by the given angle—leaving the row spacings and angles unchanged. This behavior indicates that the rows are the result of a real surface structure and not an artifact of the measurement.
artifact of the measurement.

When we observed the super-periodic structure, it frequently extended over regions greater than 1000 Å square and appeared to cover a nonnegligible fraction of the surface (perhaps 10% or more). This fraction, however, was difficult to quantify because of the relatively narrow field of view of the STM. The super-periodic structure was observed for only about 25% of the cleaves—which suggests that it may be cleavage dependent. When we observed the super-period-

 0734-211X/91/020730-05$01.00

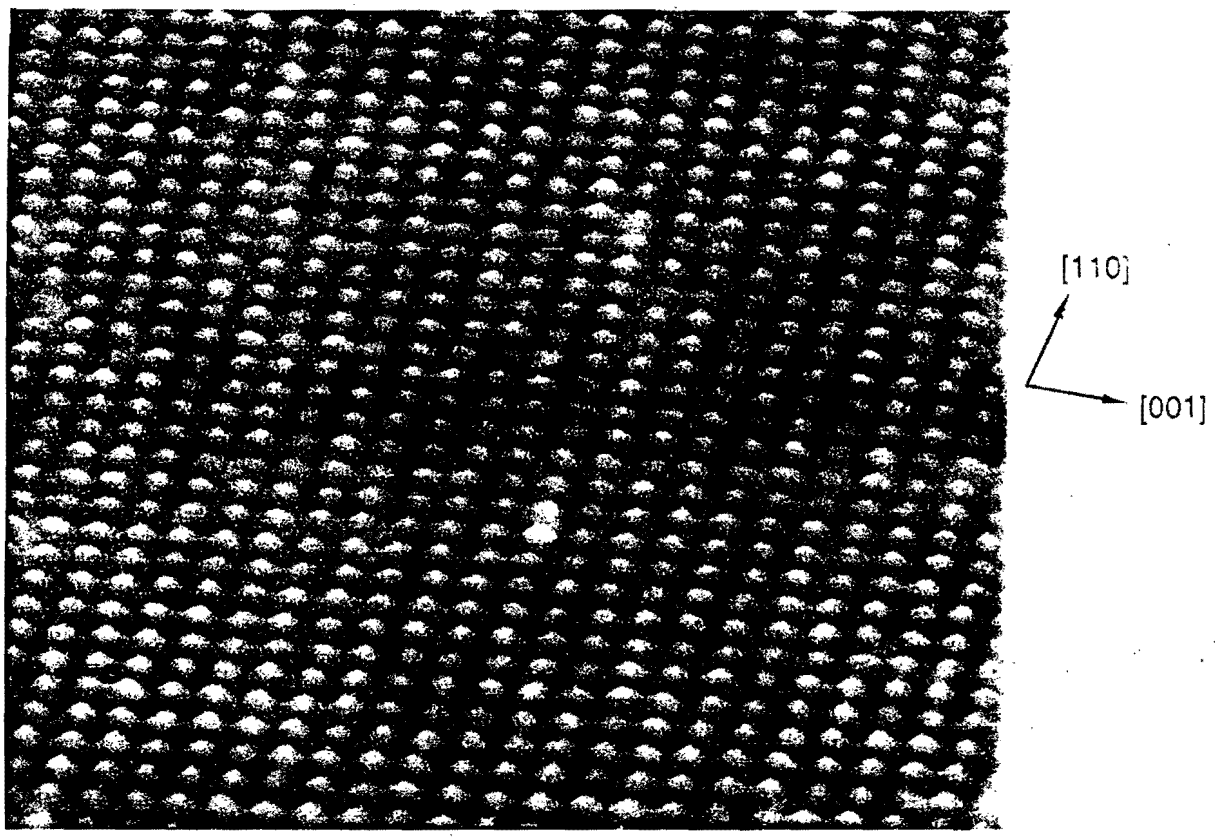


FIG. 1. STM image of the InSb(110) surface showing a large defect-free region. The surface was imaged at +0.2 V sample bias with 120 pA tunneling current. The area shown is 150×150 Å. The white features are In-derived states.

ic structure, the widely spaced rows were always parallel to the cleavage direction to within about ± 10°.

The STM results indicate that the large-distance row spacing of this super-periodic structure is 19 Å and the relative angle between these rows and the In or Sb rows is about 45°. If we assume the super-periodic structure is commensurate with the In–Sb chains, then the data are consistent with a $c(4\times6)$ reconstruction of the surface. A $c(4\times6)$ recon-

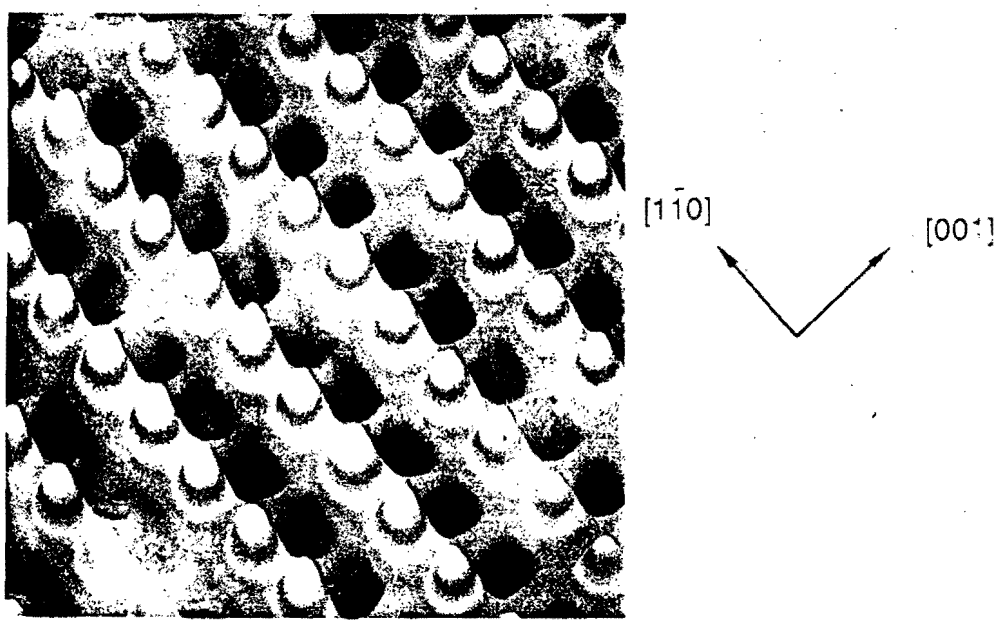


FIG. 2. STM image of the InSb(110) surface showing Sb-derived states. The sample bias was −0.2 V and the tunneling current was 100 pA. The scan range was 35×35 Å^2. The gold features are Sb derived.

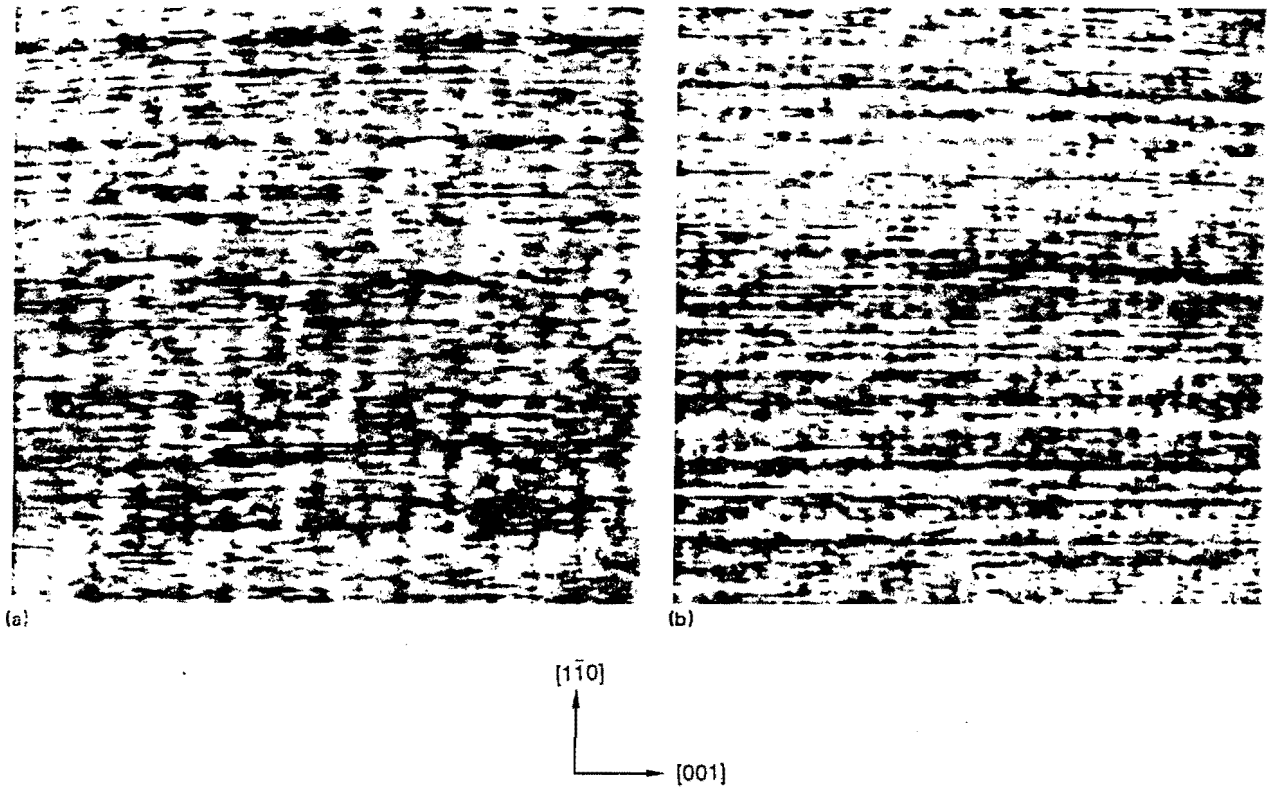


FIG. 3. (a) STM image of cleaved InSb(110) surface showing two periodicities. The vertical rows are the "normal" 1×1 In–Sb rows, the diagonal rows are the super-periodic structure with rows 19 Å apart. The range was 95×95 Å^2; the sample bias was $+0.3$ V; the scan rate was 500 Å/s. (b) The super-periodic structure was also observed under negative sample bias. The range was 190×190 Å^2; the sample bias was -0.3 V; the scan rate was 600 Å/s.

struction could produce a row structure with a row spacing of 19 Å with rows inclined 47° with respect to the In–Sb chains. A $c(4\times6)$ lattice illustrating this is shown in Fig. 4 where four atoms per lattice point are colored black. These colored atoms form rows spaced 19 Å at an angle of 47° with respect to the [1$\bar{1}$0] direction. These colored atoms may have a different electronic structure because of, perhaps, dimerization with neighboring atoms or a dimer density wave. The actual details and mechanism for the reconstruction are not fully understood at present. The STM results do suggest, however, that some regions of the InSb surface reconstruct with a $c(4\times6)$ reconstruction. Calculations are presently under way to elucidate the physics and stability of this reconstruction. Clearly additional experimentation and theoretical work are needed to further verify and understand this reconstruction.

To our knowledge this is the first observation of a new reconstruction on the cleaved InSb(110) surface. Interestingly enough, though, Lapeyre *et al.* observed cleavage dependent photoemission anomalies on cleaved GaAs(110)[7] and similar anomalies on InSb.[8] Perhaps these anomalies are associated with a cleavage-related reconstruction such as the one reported here.

The (110) surface of GaAs is one of the best understood compound semiconductor surfaces, and there is a school of thought that other zincblende semiconductors have the same surface relaxation geometry as GaAs,[9] although this latter viewpoint is certainly controversial.[10] Therefore, the observation on InSb(110) of a novel structure unknown in the physics of GaAs is unanticipated and suggests that the zincblence surfaces have a richer and more varied physics than once thought.

The image of the $c(4\times6)$ structure appears to be noisier than the 1×1 structure images. We speculate that this may be due to "softer" force constants on the metastable $c(4\times6)$ surface, and interactions between the STM tip and the surface: as the tip passes over the surface, the surface atoms move.

B. Fabrication of nanoscopic dots

Being able to modify surfaces on a nanometer scale using the tip–surface interaction is another valuable feature of STMs. Becker *et al.* demonstrated that single atoms could be manipulated on the Ge(111) surface with a single voltage pulse.[11] Eigler and Sachweizer were able to position individual Xe atoms on the Ni surface.[2] A number of authors[12] have "machined" various surfaces on the nanometer scale using various methods such as tip crashes, voltage pulses, and local heating in air, liquids, and vacuum. Here we show that nanoscopic dots can be fabricated in UHV on the InSb(110) surface.

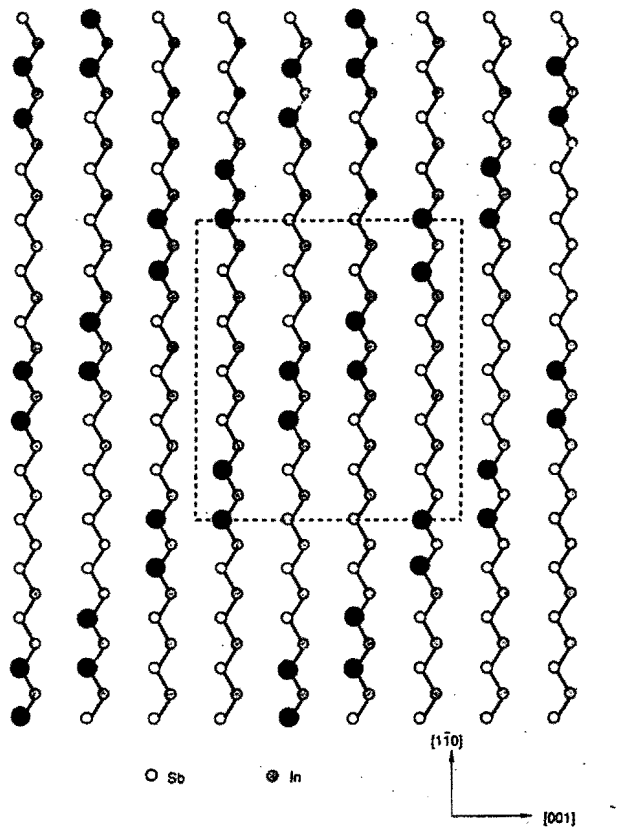


FIG. 4. A model of the $c(4\times6)$ surface structure. The small open and shaded circles are In and Sb atoms. The large black circles form rows 19 Å apart which have an angle of 43 deg with respect to the [001] direction. The dashed lines form a surface unit cell of $c(4\times6)$ symmetry.

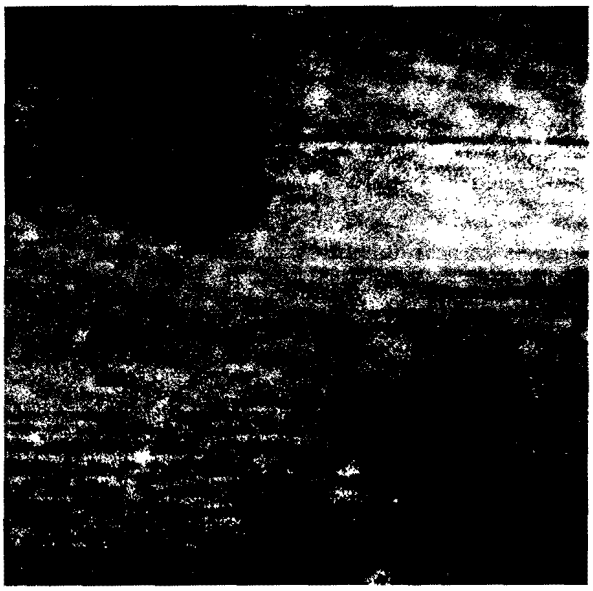


FIG. 5. Two nanoscopic dots were produced on a perfect InSb(110) surface by placing the tip over each region for about 2 min. The nanoscopic dots have radii of ~ 15 and 25 Å. Individual In-derived states can also be observed in the image. The sample was biased + 0.5 V and the tunneling current was 120 pA.

Unlike Si surfaces and other surfaces that we have studied,[12,13] the InSb(110) surface could easily be altered by the tunneling process. We found that placing the tip over the same spot on the surface for about 2 min (without scanning) frequently produced a hole or dot under the tip for small tunneling currents and voltages (tunneling current about 100 pA, basis voltage a few tenths of a volt). An example of two such holes is shown in Fig. 5. Before positioning the tip over each hole area the surface was perfect with no holes or defects. After successively positioning the tip over each hole area for about 2 min, two nanoscopic dots were formed. The resulting dots had a radius of 15–25 Å and were one to two atomic layers deep. The dots were not the result of tip crashes since there was no evidence of a tip crash in the tunneling current. Furthermore, our microscope has unusual vibrational isolation stability and we have never observed an unintentional tip crash while tunneling over flat surfaces of other materials.[13]

For writing patterns on the surface, one could use the tip as a dot-maker of a dot-matrix printer. It would be desirable, however, to be able to write on the surface while the tip is scanned. This would allow, for example, a continuous line or script pattern to be written on the surface. To this end we found that it is possible with low scan rates to produce nanoscopic holes on the InSb(110) surface while the tip is scanned. In Fig. 6 we show an example of a small dot which was formed by slowly scanning (200 Å/s) over a small 25×25 Å area of the surface. This process produced a nanoscopic dot over nearly the entire scan range after two repeated scans. The resulting dot is shown in Fig. 6 and has a radius of about 9 Å.

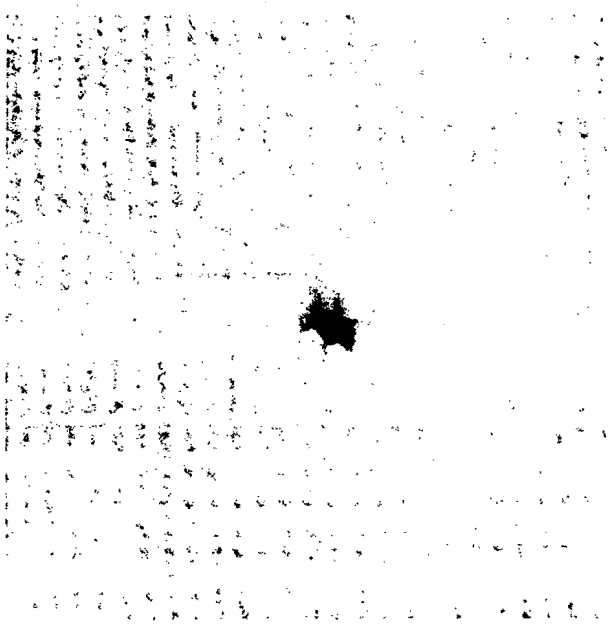


FIG. 6. A single nanoscopic dot was fabricated on the InSb(110) surface by slowly scanning (200 Å/s) the dot region. The radius of the dot is 9 Å. The sample bias was 0.2 V and the tunnel current was 100 pA.

These results demonstrate that angstrom-scale STM lithography on InSb(110) surfaces will be feasible allowing, perhaps, complicated patterns to be fabricated on this surface.

IV. CONCLUSIONS

We obtained some of the first atom-resolved STM images on the cleaved InSb(110) surface in UHV. We observed the expected 1×1 structure on this surface which is similar to that of GaAs. In addition we observed a new super-periodic structure which consisted of rows 19 Å apart rotated 47° with respect to the In–Sb chains. This structure is consistent with a $c(4\times6)$ reconstruction. Furthermore, we found that the InSb(110) surface is easily disturbed by the tunneling process. Holding the tip fixed over the same area of the surface or slowly scanning the tip over a region of the surface frequently alters the surface by producing a hole. Using this technique, we have been able to produce nanoscopic dots as small as 9 Å radius. Efforts are underway to write angstrom-scale words on the InSb(110) surface.

ACKNOWLEDGMENTS

We are grateful to the U.S. Army Research Office for their support of this work (Contract No. DAAL03-87-K-0112). We have also benefited from stimulating discussions of STM design with R. C. Jaklevic, W. J. Kaiser, S.-L. Tang, and B. S. Swartzentruber, and we thank them for their assistance in the development of our microscopy program. Our thanks also go to M.-H. Tsai and W.-M. Hu for fruitful discussions.

[1]G. Binnig, H. Rohrer, Ch. Gerber, and E. Weibel, Phys. Rev. Lett. **49**, 57 (1982).
[2]D. M. Eigler and E. K. Schweizer, Nature **344**, 524 (1990), and references therein.
[3]L. J. Whitman, J. A. Stroscio, R. A. Dragoset, and R. J. Celotta, Bull. Am. Phys. Soc. **35**, 226 (1990). L. J. Whitman, J. A. Stroscio, R. A. Dragoset, and R. J. Celotta, J. Vac. Sci. Technol. B **9**, 770 (1991); Bull. STM'90 and NANO I, 175 (1990).
[4]Y. Liang, W.-M. Hu, W. E. Packard, and J. D. Dow, Bull. Am. Phys. Soc. **35**, 227 (1990).
[5]R. M. Feenstra, J. A. Stroscio, J. Tersoff, and A. P. Fein, Phys. Rev. Lett. **58**, 1192 (1987).
[6]Pachyderm Scientific Industries Technical Report, TR1000-01, P. O. Box 3102, Tempe, AZ 85280-3102.
[7]F. Cerrina, T. R. Myron, and G. J. Lapeyre, Phys. Rev. B **29**, 1798 (1984).
[8]G. J. Lapeyre (private communication).
[9]C. Malhiet, C. B. Duke, and D. J. Chadi, Surf. Sci. **149**, 366 (1985).
[10]M.-H. Tsai, R. V. Kasowski, and J. D. Dow, Phys. Rev. B **40**, 9818 (1989); M.-H. Tsai, J. D. Dow, K.-P. Wang, and R. V. Kasowski, Superlatt. Microstruct. **6**, 431 (1989).
[11]R. S. Becker, J. A. Golovchenko, and B. S. Swartzentruber, Nature **325**, 419 (1987).
[12]W. E. Packard, Y. Liang, N. Dai, J. D. Dow, R. Nicolaides, R. C. Jaklevic, and W. J. Kaiser, J. Microsc. **152**, 715 (1988), and references therein.
[13]W. E. Packard, N. Dai, J. D. Dow, R. C. Jaklevic, W. J. Kaiser, and S. L. Tang, J. Vac. Sci. Technol. A **8**, 3512 (1990).